

Discovery and development of novel inhibitors for the kinase Pim-1 and G-Protein Coupled Receptor Smoothened

Dissertation

zur

**Erlangung des Doktorgrades
der Naturwissenschaften**

(Dr. rer. nat.)

dem

Fachbereich Pharmazie der
Philipps-Universität Marburg

vorgelegt von

Corey Taylor

aus

Adelaide, South Australia

Marburg/Lahn 2019

Erstgutachter

Prof. Dr. Peter Kolb
Institut für Pharmazeutische Chemie
Philipps-Universität Marburg

Zweitgutachter

Prof. Dr. Gerhard Klebe
Institut für Pharmazeutische Chemie
Philipps-Universität Marburg

Eingereicht 09.04.2019

Tag der mündlichen Prüfung am 06.06.2019

Hochschulkennziffer: 1180

Contents

1	Introduction	11
1.1	Abstract	13
1.2	Zusammenfassung	15
1.3	Acknowledgements	17
2	Overview: techniques and approaches	21
2.1	Structure-based virtual screening	23
2.2	Fragment-based drug-discovery	29
2.3	Molecular docking	31
2.3.1	Physics-based methods	31
2.3.2	Shape-based methods	34
2.4	Molecular dynamics simulations	35
2.4.1	Equilibrium simulation	36
2.5	Energy minimisation	40
2.5.1	Steepest descent minimisation	41
2.5.2	Conjugate gradient minimisation	42
2.5.3	Minimisation and ligand entropy	43
2.6	Ligand-based techniques	46
2.6.1	Similarity and fingerprinting	47
3	Novel ligands for the kinase Proviral Integration site for MuLV (Pim-1)	49
3.1	Introduction	51
3.1.1	The kinase Pim-1	51
3.1.2	Structure-based Virtual Screening	55
3.2	Materials and methods	57
3.2.1	Protein preparation	57
3.2.2	Ligand preparation	58

3.2.3	Molecular dynamics simulations	60
3.2.4	Docking and minimisation	63
3.3	Results	66
3.3.1	Project 1: Docking software comparisons vs structures from MD simulation	66
3.3.2	Project 2: synthetically-feasible extensions of an internal fragment library	68
3.3.3	Project 3: novel inhibitors from the ZINC dataset	75
3.3.4	Project 4: extensions of truncated fragment f200	81
3.4	Discussion and conclusions	84
4	Novel ligands for the Smoothened receptor	97
4.1	Introduction	99
4.2	Material and methods	103
4.2.1	Receptor preparation	103
4.2.2	Ligand preparation	107
4.2.3	Docking and pose minimisation	108
4.2.4	qPCR assay	109
4.3	Results	109
4.3.1	Identifying the region of Fo-Fc density and mapping SMO's allosteric pocket	109
4.3.2	ZINC virtual screen - Round 1	116
4.3.3	Assay results - Round 1	121
4.3.4	Minimisation of the binding pocket found more SMO poses	121
4.3.5	ZINC virtual screen - round 2	127
4.3.6	Assay results - Round 2	131
4.4	Discussion and conclusions	131
5	Perspective	141
5.1	Introduction	143
5.2	What was learned from this thesis	144
5.2.1	Chemically diverse synthetically feasible ligands were found against Pim-1	144
5.2.2	Novel ligands were also found in focussed libraries for diverse targets .	145
5.2.3	Structural information matters	146

5.2.4	One should always be prepared for surprises	147
5.2.5	Some aspects didn't work as well as we would like	149
5.2.6	What is a 'hit'?	150
6	Bibliography	153
7	Appendix	169
7.1	Novel ligands against the kinase Proviral Integration site for MuLV (Pim-1) .	171
7.1.1	Project 1: Docking software comparisons vs structures from MD simulation	171
7.1.2	Project 2: synthetically-feasible extensions of an internal fragment library	177
7.1.3	Project 3: novel inhibitors from the ZINC dataset	185
7.1.4	Project 4: extensions of truncated fragment f200	186
7.2	Novel inhibitors against the Smoothened receptor	189
8	Definitions	203
9	Curriculum vitae	207

List of Figures

2.1	UCSF DOCK receptor:ligand sphere matching	33
2.2	FRED docking process	35
2.3	The ECFP and unique identifier generation process	48
2.4	Generation of bit-strings from unique identifiers for the ECFP generation process	48
3.1	PIM in the JAK2 pathway	52
3.2	Pim-1 basic kinase structure	53
3.3	Pim-1 in complex with AMPPNP and non-conserved regions	54
3.4	Pim-1 X-ray structures, published ligands with short-chain linkers	55
3.5	Pim-1 binding pocket with aniline triazole scaffold and predicted interactions	59
3.6	Pim-1 binding pocket with quinoxaline scaffold and predicted interactions . .	60
3.7	Example Pim-1 apo MD snapshot with 20% ethanol/water solvent mixture . .	62
3.8	Z'LYTE biochemical assay schematic	65
3.9	Schematic of Project 2 preparation, docking and post-processing	70
3.10	Predicted binding mode of 29 against Pim-1	71
3.11	Per-cent phosphorylation against ligand concentration of tested compounds displaying activity against Pim-1	71
3.12	Predicted binding mode of the (E)-isomer of 26 against Pim-1	73
3.13	Per-cent phosphorylation against ligand concentration of 52 and Staurosporine against Pim-1	74
3.14	Predicted binding mode of 52 overlayed with extended SCUBIDOO fragment 4012414	74
3.15	Predicted binding mode of 59 with Pim-1	76
3.16	X-ray structure of 61 with Pimtide (left) and docking prediction (right) against Pim-1	77
3.17	X-ray structure of 57 with Pimtide (left) and docking prediction (right) against Pim-1	78

3.18	Ligands 57 (left) and 61 (right) overlayed with docking predictions in complex with Pim-1 and Pimtide	79
3.19	Per-cent phosphorylation against ligand concentration of compounds 57 , 58 and 61 against Pim-1	79
3.20	Compound 58 docking prediction against Pim-1	81
3.21	Areas of solvent density below $-1.0 \text{ kcal}\cdot\text{mol}^{-1}$ overlayed with crystollagraphic ligand and water molecules	82
3.22	Positions and distances between ligand and buried waters for f200 amide and iminol tautomers	83
3.23	Predicted binding modes against Pim-1 of molecules 62 and 68	84
3.24	Differences between the positions of K67 and F49 (P-Loop) in 3BGP and apo_MD structures	86
3.25	Suggested workflow for various docking codes when coupled with molecular dynamics simulation	88
3.26	Docking predictions of 26 vs 62	90
3.27	Crystal structure of 61 in complex with Pimtide	93
3.28	Other targeting strategies for Pim-1	94
4.1	SMO structural features	100
4.2	Hedgehog activated and inactivated pathways in cells	101
4.3	Published drugs and inhibitors against SMO	102
4.4	Region of Fo-Fc density with from unpublished X-ray structure with aligned crystal structure of Vismodegib	105
4.5	Indicative sections of the CURED	110
4.6	Example A-carboline docking pose within the CURED	111
4.7	Distance to E481 and Q477 from the CURED	112
4.8	Distance to D384 from the CURED	113
4.9	Docked methanol poses within the CURED	113
4.10	Favourable N-methylpyrrole vs unfavourable benzene in similar position within the CURED	114
4.11	Unfavourable indole poses within the CURED	115
4.12	R400 position in CURED X-ray structure vs position with respect to Vismodegib	115
4.13	Example molecule indicative of moieties favourable in the CURED region . .	116
4.14	Predicted binding mode of K5 against SMO	118
4.15	Predicted binding mode of K4 against SMO	119

4.16	Normalised Gli mRNA levels for K1 - K11 at 10 μ M and 100 μ M	121
4.17	Example molecular pose with unlikely pyrimidine protonation state and primary amide with torsional strain	122
4.18	Vismodegiband residues within 8 Å of it in ECL1, 3 and 4 and TM6	123
4.19	Selected minimised binding pocket residues comparing crystal structure and after minimisation with CHARMM and MOE	124
4.20	Residues that were tarted below H470	126
4.21	Predicted binding mode of compound K2.10 against SMO	128
4.22	Predicted binding mode of compound K2.19 against SMO	129
4.23	Normalised Gli mRNA levels for K2.1 - K2.20 at 10 μ M and 100 μ M	131
4.24	Other work involving oxysterol inhibitors to SMO	133
4.25	Predicted binding modes of K5 and K6	135
4.26	Predicted binding pose of K2.7 , K2.11 and K2.12	136
4.27	Interaction matrix of all compounds tested in Rounds 1 and 2	137
7.1	Prepared Pim-1 docking binding site - SEED	171
7.2	Prepared Pim-1 docking binding site - FRED	172
7.3	Prepared Pim-1 docking binding site - DOCK	172
7.4	Pim-1 Thermal Shift Assay (TSA) results	173
7.5	IC ₅₀ curves for compounds 57 and 58 - Eurofins	186
7.6	Mean distances (Å) between simulated amide and iminol tautomers	186
7.7	Mean RMSD (Å) between simulated amide (red trace) and iminol (black trace) tautomers	187
7.8	Prepared SMO docking binding site - SEED	190
7.9	Prepared SMO docking binding site - unpublished structure	191
7.10	Prepared SMO docking binding site - 5L7I	192
7.11	Prepared SMO docking binding site - 4N4W	193
7.12	ROC plots of molecular enrichment of known actives for prepared structures .	197

List of Tables

2.1	Various classical thermodynamic ensemble types and properties	37
3.1	Comparison mean and standard deviations of RMSDs (Å) from X-ray structures	68
3.2	All compounds tested from extensions via <i>in silico</i> reactions with the internal library, their measured K_i and 2D structure	72
3.3	Predicted Pim-1 inhibitors, sample mean and standard deviation of temperature shifts in thermal-shift assays conducted, IC_{50} as determined by third-party vendor, measured K_i 's and 2D structures of all ligands	80
4.1	Predicted Round 1 SMO inhibitors	120
4.2	Descriptive statistics of differences in ranking and score for 440 molecules common between ZINC12 dockings	125
4.3	Descriptive statistics of differences in ranking and score for the 1699 molecules common between ZINC12 dockings	127
4.4	Predicted Round 2 SMO inhibitors	130
7.1	Mean RMSDs (Å) of Pim-1 apo structure from MD simulations - Project 1 . .	173
7.2	RMSD (Å) from the crystal structure of fragments from the internal fragment library - 3BGP	174
7.3	RMSD (Å) from the crystal structure of fragments from the internal fragment library - apo_MD	175
7.4	RMSD (Å) from the crystal structure of fragments from the internal fragment library - 3BGP_MD	176
7.5	Number of derivatives from fragments created <i>in silico</i> - Project 2	177
7.6	Complete list of predicted Pim-1 inhibitors, K_i of all that were tested and 2D structure - Project 2	183
7.7	Selected Szybki minimisation output for synthesised inhibitors 26 , 27 and 29 - Project 2	183

7.8	Selected Szybki minimisation output for synthesised inhibitors 30 , 39 and 40	
	- Project 2	184
7.9	Selected Szybki minimisation output for synthesised inhibitors 42 , 47 and 48	
	- Project 2	184
7.10	Predicted Pim-1 inhibitors, aniline triazole series - Project 3	185
7.11	Mean % residual activity at 10 μ M of Pim-1 inhibitors - Eurofins	185
7.12	Calculated IC ₅₀ 's of Pim-1 inhibitors - Eurofins	186
7.13	Predicted quinoxaline dervative inhibitors - Project 4	188
7.14	Exhaustive list of molecules ruled out by collaborators as fitting within the CURED	189
7.15	Protonation states of Histidines in all Smoothened structures	189
7.16	All molecular 'probes' used in docking to the unpublished structure	194
7.17	Calculated interaction energies in kcal.mol ⁻¹ , probes outside blob	195
7.18	Calculated interaction energies in kcal.mol ⁻¹ , probes inside blob	196
7.19	Smoothened residues within 8Å of the Vismodegib binding position by helix/loop	198
7.20	Descriptive statistics for distance in RMSD of selected residues from X-ray structure of PDB:5L7I of CHARMM-minimised and MOE-minimised receptor structures	198
7.21	Calculated receptor energies of receptors after minimisation with CHARMM and MOE	199
7.22	Tarted Smoothened residues	199
7.23	Molecules known to be active against SMO that scored high when docked to the tarted receptor but poorly against the untarted receptor	200
7.24	Descriptive statistics of ECFP4 Tanimoto similarity of Round 2 molecules with Lacroix et al	201
7.25	Mean estimated energies from SZYBKI (kcal.mol ⁻¹) for ZINC12 Round 1 vs Round 2 molecules	201
7.26	Mean estimated energies from SZYBKI (kcal.mol ⁻¹) for Round 2 molecules - ZINC12 vs ZINC15	201

Chapter 1

Introduction

1.1 Abstract

Investigation of the cause of disease is no easy business. This is particularly so when one reflects upon the lessons taught us in antiquity. Prior to the beginning of the last century, diagnosis and treatment of diseases such as cancers was so bereft of hope that there was little physicians could offer in the way of comfort, let alone treatment. One of the major insights from investigations into cancers this century has been that those involved in research leading to treatments are not dealing with a singular malady but multiple families of diseases with different mechanisms and modes of action. Therefore, despite the end game being similar in cancers, that of uncontrolled growth and replication leading to cellular dysfunction, different diseases require different approaches in targeting them.

This leads us to a particular broad treatment approach, that of drug design. A drug is, in the classical sense, a small molecule that, upon introduction into the body, interacts with biochemical targets to induce a wider biological effect, ideally with both an intended target and intended effect. The conceptual basis underpinning this ‘lock-and-key’ paradigm was elucidated over a century ago and the primary occupation of those involved in biochemical research has been to determine as much information as possible about both of these protein locks and drug keys. And, as inferred from the paradigm, molecular shape is all-important in determining and controlling action against the most important locks with the most potent and specific keys.

The two most important target classes in drug discovery for some time have been protein kinases and G Protein-Coupled Receptors (GPCRs). Both classes of proteins are large families that perform very different tasks within the body. Kinases activate and inactivate many cellular processes by catalysing the transfer of a phosphate group from Adenosine Tri-Phosphate (ATP) to other targets. GPCRs perform the job of interacting with chemical signals and communicating them into a biological response. Dysfunction in both types of proteins in certain cells can lead to a loss of biological control and, ultimately, a cancer.

Both of kinases and GPCRs have entirely different chemical structures so structural knowledge therefore becomes crucial in any approach targeting cells where dysfunction has occurred. Thus, for this thesis, a member from each class was investigated using a combination of structural approaches. From the kinase class, the kinase Proviral Integration site for MuLV (Pim-1) and from the GPCR class, the cell membrane-bound Smoothed receptor (SMO).

The kinase Pim-1 was the target of various approaches in chapter 3. Although a heavily studied target from the mid-2000's, there is a paucity of inhibitors targeting residues more remote from structural characteristics that define kinases. Further limiting extension possibilities is that Pim-1 is constitutively active so no inhibitors targeting an inactive state are possible.

An initial project (Project 1) used the known binding properties of small molecules, or, 'fragments' to elucidate structural and dynamic information useful for targeting Pim-1. This was followed by three projects, all with the goal of inhibitor discovery, all with different foci. In Project 2, fragment binding modes from Project 1 provided the basis for the extension and development of drug-like inhibitors with a focus on synthetic feasibility. In contrast, inhibitors were found in Project 3 via a large-scale public dataset of purchasable molecules that possess drug-like properties. Finally, Project 4 took the truncated form of a particularly attractive fragment from Project 1 that was crystallised with Pim-1, verified its binding mode and then generated extensions with, again, a focus on synthetic feasibility.

The GPCR SMO has fewer molecular studies and much about its structural behaviour remains unknown. As the most 'druggable' protein in the Hedgehog pathway, structural studies have primarily focussed on stabilising its inactive state to prevent signal transduction. Allied to this is that there are generally few inhibitors for SMO and the drugs for cancers related to its dysfunction are vulnerable to mutations that significantly reduce their effectiveness or abrogate it entirely. The elucidation of structural information is therefore of high priority.

An initial study attempting to identify an unknown molecule from prior experiments led to insights regarding binding characteristics of specific moieties. This was particularly important to understand not just where favourable moieties bind but also sections of the SMO binding pocket with unfavourable binding. In both subsequent virtual screens performed in Chapter 4, the primary aim was to find new drug-like inhibitors of SMO using large public datasets of commercially-available molecules. The initial screen retrieved relatively few inhibitors so the binding pocket was modified to find a structural state more amenable to small molecule binding. These modifications led to a significant number of new, chemically novel inhibitors for SMO, some structural information useful for future inhibitors and the elucidation of structure-activity relationships useful for inhibitor design. This underpins the idea that structural information is of critical importance in the discovery and design of molecular inhibitors.

1.2 Zusammenfassung

Die Untersuchung der Ursache einer Krankheit ist keine leichte Aufgabe. Dies gilt insbesondere, wenn man über die Lehren nachdenkt, die wir aus der Antike gelernt haben. Vor Beginn des letzten Jahrhunderts war die Diagnose und Behandlung von Krankheiten wie Krebserkrankungen so hoffnungslos, dass Ärzte kaum Trost, geschweige denn eine Behandlung anbieten konnten. Eine der wichtigsten Erkenntnisse aus der Erforschung von Krebserkrankungen in diesem Jahrhundert ist, dass es sich dabei nicht um eine einzelne Krankheit, sondern um mehrere Familien von Krankheiten mit unterschiedlichen Mechanismen und Wirkungsweisen handelt. Trotz der Tatsache, dass der finale Verlauf von Krebserkrankungen ähnlich ist, d.h. unkontrolliertes Wachstum und Replikation, die zu zellulärer Dysfunktion führen, erfordern die verschiedenen Krankheiten daher unterschiedliche Ansätze in der Behandlung.

Dies führt uns zu einem besonders breiten Behandlungsansatz, dem des Wirkstoffdesigns. Die beiden wichtigsten Zielklassen in der Wirkstoffforschung sind seit längerem Proteinkinasen und G-Protein gekoppelte Rezeptoren (GPCRs). Beide Kategorien von Proteinen umfassen große Familien, die sehr unterschiedliche Aufgaben im Körper erfüllen. Dysfunktionen beider Arten von Proteinen in bestimmten Zellen können zu einem Verlust der biologischen Kontrolle und letztlich zu Krebs führen. Sowohl Kinasen als auch GPCRs besitzen völlig unterschiedliche chemische Strukturen, sodass bei der Adressierung dysfunktionaler Zellen strukturelles Wissen von entscheidender Bedeutung ist. Daher wurde in dieser Arbeit je ein Mitglied aus beiden Klassen mit einer Kombination verschiedener strukturbasierter Ansätze untersucht. Aus der Familie der Kinasen die Kinase Proviral Integration Site for MuLV (Pim-1) und aus der Familie der GPCRs der zellmembrangebundene Smoothed Rezeptor (SMO).

Die Kinase Pim-1 war das Ziel verschiedener Ansätze in Kapitel 3. Obwohl es sich um ein stark untersuchtes Target aus der Mitte der 2000er Jahre handelt, gibt es nur wenige Inhibitoren, die Aminosäuren adressieren, welche weiter von strukturellen Merkmalen entfernt sind, welche Kinasen ausmachen. Weitere Einschränkungen sind dadurch gegeben, dass Pim-1 konstitutiv aktiv ist, sodass bei der Entwicklung von Inhibitoren nicht auf eine inaktive Proteinkonformation abgezielt werden kann.

Ein erstes Projekt (Projekt 1) nutzte bekannte Bindungseigenschaften kleiner Moleküle, um strukturelle und dynamische Informationen zu gewinnen, die bei der Adressierung von Pim-1

nützlich sind. Es folgten drei Projekte mit dem Ziel, Inhibitoren unter Berücksichtigung unterschiedlicher Gesichtspunkte zu entwickeln. In Projekt 2 bildeten Fragmentbindungsmodi aus Projekt 1 die Grundlage für die Erweiterung und Entwicklung von medikamentähnlichen Inhibitoren mit einem Schwerpunkt auf synthetischer Darstellbarkeit. In Projekt 3 wurden Inhibitoren aus einem großen, öffentlich zugänglichen Datensatz von kaufbaren Molekülen mit medikamentenähnlichen Eigenschaften herausgesucht. Schließlich wurde in Projekt 4 die verkürzte Form eines besonders attraktiven Fragments, das mit Pim-1 kristallisiert wurde, aus Projekt 1 aufgenommen, der Bindungsmodus verifiziert und dann Erweiterungen generiert, wiederum mit dem Fokus auf der synthetischen Darstellbarkeit.

Über den GPCR SMO gibt es weniger molekulare Studien und vieles über sein strukturelles Verhalten ist weiterhin unbekannt. Als das am besten adressierbare Protein im Hedgehog Signalweg haben sich bisherige strukturbasierte Studien in erster Linie darauf konzentriert, den inaktiven Zustand zu stabilisieren, um so eine Signaltransduktion zu verhindern. Daher gibt es generell nur wenige Inhibitoren für SMO und die Medikamente für Krebserkrankungen im Zusammenhang mit der Dysfunktion von SMO sind anfällig für Mutationen, was ihre Wirksamkeit signifikant verringern oder ganz aufheben kann.

Eine erste Studie, in der versucht wurde, ein unbekanntes Molekül aus früheren Experimenten zu identifizieren, resultierte in Erkenntnissen über die Bindungseigenschaften bestimmter Struktureinheiten. Dies war besonders wichtig, um nicht nur Bereiche günstiger Interaktionen, sondern auch Bereiche ungünstiger Interaktionen in der SMO-Bindungstasche zu verstehen. In den beiden nachfolgenden virtuellen Screens, die in Kapitel 4 beschreiben werden, bestand das Hauptziel darin, in großen, öffentlich zugänglichen Datensätzen kommerziell verfügbarer Moleküle neue medikamentenähnliche Inhibitoren für SMO zu finden. Der erste Screen sowie die nachfolgenden Modifikationen resultierten in einer signifikanten Anzahl chemisch neuartiger Inhibitoren für SMO, sowie strukturellen, für die Entwicklung zukünftiger Inhibitoren nützlichen Erkenntnissen und der für das Inhibitor-Design nützlichen Aufklärung von Struktur-Aktivitäts-Beziehungen. Dies untermauert die Idee, dass Strukturinformationen für die Entdeckung und das Design molekularer Inhibitoren von entscheidender Bedeutung sind.

1.3 Acknowledgements

Where there is no echo there is no description of space or love. There is only silence.”

MZ Danielewski - House of Leaves

Research, as in life, is a team sport. Everything I learned about morality, bravery, fair-play and duty I certainly learned from sport but also from being part of strong teams generally. It has been the most consistent theme in my CV. In my view, the first level of building a reasonable person is to acknowledge that one is not alone. And with that comes a responsibility to accept that any success owes many others.

However, to really get things done, one must retain a sense of the individual. The isolation of being an individual, replete with standards and moral boundaries, and working within a team environment, you have to develop a sense of direction, a pull from the chest. Any effort is imbued with colours by the individual that will make other individuals, within and without the team, care about it. The delicate and difficult balancing act between those two worlds, I feel, takes a lifetime to understand and put into action.

It was this pull that led me toward doing a PhD. It was always something in the extremities of my mind but before it, there were other things to do. The pull certainly led me away from other possibly more lucrative careers and decided for me that moving continents was the way forward to extract as much as possible from a rare opportunity in life. That being to take ones intellectual growth seriously.

My time within the Kolb Lab, Marburg and Europe generally has been a period of extraordinary change for me personally. Not all changes are visible. But as I had been away from science and chemistry for over a decade, the first steps were merely to become familiar with where the fields and subfields I thought I knew were at. I was relieved at how much was familiar but it was very quickly apparent just how much catching up I also had to do. To that end, I wish to acknowledge my team members and how much they helped me along the way because I feel fairly certain they're not fully aware of their contribution.

I would like to take the opportunity to thank the most influential persons on my professional life. **Narmon Tulsi**, a man with whom I have collaborated with professionally and played many gigs with, taught me that it really is quite possible to be friendly, loved and absurd within any working context but that one really should save all your Sent Items just in case (trust, *but verify*). I love you, man. **Tauto Sansbury**, of course he's really family.

Indescribable bravery but an object lesson that real bravery is, in the face of The Struggle that kills other, being kind. *Illegitimi non carborundum*. **Dr. Sophie Pointer**, you started this and you're an absolute badass. Cheerios to **Sergeant Peter Broderick** who completely upended what I thought I knew about carving out one's own opportunities where others would have seen none, **Det. Sen. Constable Darren Crowhurst** that it's both possible and perfectly okay to show kindness in the job (also trust/verify) and **Snr. Sgt. Paul Lewandowski** - uneasy lies the head that wears a crown, particularly in policing environment, but thank you for utterly modelling integrity and sensitivity. Thank you to the, sadly, recently deceased **Chris Gascoigne** for being a huge part of my escape plan toward research. Finally, I'd like to thank **Mrs Katherine Economides** for proving that one can be vulnerable, a leader and a friend.

In terms of research, it's quite difficult to overstate the influence of **Prof. David Roder** and my Master supervisors **Dr Richard Leibbrandt** and **Prof. David Powers**. Roder for proving that you don't have to be crazy to do research but it helps, David and Richard for 'jamming' with me on ideas, encouraging total free thinking and insane possibilities. All of my Pim-1 collaborators for fancy ideas and shared frustrations in tackling this beast; **Christof, Lukas** and **Matthias** - thank you ballboys, thank you linesmen. Particularly I'd like to thank **Prof. Dr. Wibke Diederich** for her kindness, supernatural insight and the way she says "Cooooo!". Thank you to the Smooth(ened) crew for doing some heroic work on a pretty rough target in terms of lab work; **Professor. Dr. Christian Siebold, Eamon** for some lovely emails and **Rebekkah** for nearly unnatural levels of enthusiasm.

The wider Klebe lab always had a fun vibe so I would like to acknowledge that it has been full of good people during the course of my PhD. First of all, thank you to the engine room, the people who really keep things working, **Christian** and **Lydia**, the latter particularly patient and caring with the numerous mistakes I made. Also I would like to say thank you to **Babsie** for possessing the loudest laugh in recorded history, **Felix** for reminding me to turn it off and on again and **Jonathan** for some sweet jams and generalised guitar nerdery. Last, but certainly not least, a huge thanks to **Tobi** for always valuable input, scientific help and just generally being a man one can really rely on. That jam in NYC is still on, by the way, so take yer damn bass.

The working environment of the Kolb Lab has been fabulous and supportive and that, of course, is reflective of the people contained within it. Firstly, **Denis** and **Florent** for extraordinary patience and hilarity in my new room and life in my first year as a PhD

student, especially when it seemed Denis knew every new nerd thing ahead of time (how did you Kung Fury before me, maaaaaan). Thank you to **Gio** for introducing various levels of disaster into science (fail, crush and baaaad), **Anja** one of the most utterly charming, sweetest people I've met who also happens to be a laser-sharp mind, **Goodboye Drabek** for finding the most inappropriate moments to giggle with breathtaking consistency, **Johanna** (no nickname, I wouldn't dare) and **La Donna Mafia** - all my respects. Finally, to **Spicy**, for making every nice moment extremely awkward, for laughs that will echo throughout eternity annoying everyone who hears and for always, always finding a way to argue even when in a hopeless position. All I ask is for you to please shut up. Just this once. Please? Will ya? Huh?

I would like to express my sincere appreciation for the participation of **Prof. Dr. Moritz Bünemann** and **Prof. Dr. Cornelia Keck** on my thesis panel. I would also like to express my sincere thanks to **Prof. Dr. Gerhard Klebe**. As my co-supervisor, he was not intimately involved in day-to-day supervision, of course. However, in every meeting he participated in or every seminar I gave where he asked questions, he communicated a deep knowledge and care for the scientist and the person. It was a moment of growth on every occasion thinking about questions posed and the scientific implications. So for that, I thank him for the lessons and doing it in a manner that communicates nurturing and breeds leaders.

But most of all, I offer the most inadequate of thanks to my supervisor **Heisenberg Prof. Dr. Peter Kolb**. Only with the passing of time have I been able to more fully recognise and appreciate the decision-making, balancing of risks, care and attention that goes into keeping a burgeoning research group ticking over smoothly and stable. I really knew I'd like this guy when he earnestly presented me a schedule to meet all lab members prior to interviewing. Despite the stresses and difficulties of being in charge, Peter has been available and attentive in the first instance but then creative, supportive, often cheery and unfailingly positive with ideas and difficulties. On top of all of that, he is one of those very rare people, a listener who has two ears, one mouth and uses them in proportion. A diamond man, thoroughly decent, I know I'll be counting the lessons for years. All in the lab should be praying to whatever personal \$DEITY they worship that they have him as their supervisor.

One cannot begin to quantify the impact of family on one's life. I can't even assess it. To **Mum, Dad**, my sisters **Sandy** and **Misty**, you're the reasons I'm here and nothing I can say will be enough.

Chapter 2

Overview: techniques and approaches

2.1 Structure-based virtual screening

Early-stage drug discovery efforts most commonly begin with the identification of chemical compounds displaying pharmacological activity against a target protein. This is broadly achieved via two approaches: High-Throughput Screening (HTS), where large libraries of molecules (10^5 to 10^6 /day) are physically tested against a protein, and Structure-Based Virtual Screening or Structure-Based Drug Design (SBVS/SBDD), where molecules are evaluated computationally prior to synthesis and physical testing.

The processes involved in HTS have been optimised over many years with attendant miniaturisation of assay plates and automation amongst the first wave of innovation [1]. There were other innovations such as in combinatorial synthesis, where structurally diverse arrays of small molecules, or, ‘building blocks’, are combined to generate huge chemical libraries [2] and in assay methods [3] but HTS campaigns are generally characterised by a low ratio of signal-to-noise. The efficiencies gained by SBVS are intended to solve the inefficiencies of HTS, namely, high cost in terms of both time and money but, primarily, the often indiscriminate or imprecise investigation of a molecular target.

SBVS takes as its basis a foundational concept within molecular biology that from molecular form flows biological function. Operationalised, this infers knowing the molecular structure of a biological macromolecule, such as a protein receptor, means understanding and predicting its function. Experimental methods (e.g. X-ray and NMR crystallography) have proven wildly successful at cataloguing receptor structures to the atomic level, have provided a basis for the beginnings of computational methods, that is, a 3-dimensional model of receptor structures and for establishing activity against receptors.

However these are often prohibitively expensive methods in terms of time and labour so do not lend themselves to large-scale testing of molecular libraries. Instead, they are often used in an incremental manner to establish the binding mode of molecules known to bind and then with derivatives. Cheaper methods such as molecular binding assays [4] (e.g. fluorescence anisotropy) and signal transduction assays enable larger-scale testing of compounds against biological targets.

Compounds tested in these assays can be broadly categorised according to purpose. Small molecules, or, ‘fragments’, are low molecular weight (≤ 250 Da) compounds that are biologically active. These can be as small as a ring system (e.g. indole) and are often also termed ‘scaffolds’ as they tend to be a generic molecular entity that can be built-upon. Molecular

frameworks apparently capable of serving as ligands for multiple targets led to the development of ‘privileged scaffolds’. These are molecular entities, such as a quinoline, that tend to feature in many drug molecules (e.g. Amodiaquin, Camptothecin) where libraries including these entities are synthesised to target broad classes of targets [5].

‘Lead-like’ molecules are larger ($\sim 300\text{--}400\text{Da}$) molecules demonstrated or believed to be therapeutically useful but with scope for further optimisation to improve potency. Although classes of approved drugs do exist that are no bigger than lead-like molecules [6], generally ‘drug-like’ compounds are larger ($\leq 500\text{Da}$), more potent and possess properties (e.g. $\log P$, polar surface area, bioavailability) that improve the drug’s ability to reach targeted parts of the body and reduce undesirable reactions such as toxicity and carcinogenicity [7] [8].

That these molecules are already known to be drug or lead-like molecules and have already been synthesised potentially ensures a lack of novelty is built into the process. Thus, the emergence of computer-based methods to more efficiently develop inhibitors with new chemistry. In one such approach, Virtual Screening (VS), databases of commercially-available molecules firstly have 3D representations elucidated from a source such as X-ray crystallography and are tested via computational methods against a known protein target, also typically sourced from X-ray crystallography. Predictions of a given molecule’s binding pose and affinity are made and then tested experimentally.

The size and characteristics of molecular databases are concerns [9]. The size of chemical space is vast but large portions of it are biologically uninteresting. So, whilst screening larger databases would be both slow and inefficient, strategies targeting biological relevancy are not without risk or compromise. Many biologically-relevant small-molecule databases have been made available online with library sizes in the thousands (e.g. WOMBAT [10], KEGG [11]) millions (e.g. ChEMBL [12], PubChem [13]) and hundreds of millions (ZINC [14]).

Issues remain in handling even mid-sized databases in the range of hundreds of Gb/Tb within a reasonable timeframe without requiring specialised computing facilities, such as a High-Performance Computing Cluster (HPCC). Reducing the computing effort often requires decisions prior to calculations being performed (e.g. only target molecules of lead/fragment size, $\log P$ cut-offs) or in finding ‘representative’ molecules afterward (e.g. clustering methods) that are target-specific and require expert knowledge of the target itself.

The *de novo* molecule design approach allows the computational chemist to instead design much smaller databases of molecular structures based upon the 3D structure of the receptor’s binding pocket. This essentially guides a progressive optimisation of a given molecular

representation mapped to $\Delta G_{\text{binding}}$ whilst allowing for the role of water in the binding process and taking into consideration protein flexibility [15]. This, however, can and does often result in the loss of information regarding Absorption, Distribution, Metabolism, Excretion and Toxicity (ADMET) that is internalised by curated databases of molecules known to be biologically active. Prioritising the optimisation of binding properties and *in vitro* efficacy, more often than it should, leads to molecules that fail in clinical trials ('late-stage attrition') with failure on grounds such as toxicity or adverse effects reported in up to 50% of drug candidates [16]. This has led to a conclusion that profiling of ADMET properties and omission of molecules from further consideration should occur at earlier stages, namely, at the screening stage.

Secondly, both ligands and receptors are flexible molecules and the degree of flexibility increases dramatically with the number of rotatable bonds in any given molecule so ensuring adequate conformational sampling of both is challenging. There is often difficulty in ensuring a reasonable starting structure from X-ray crystallographic studies, often stemming from high motions in the proteins under study. These motions often cause large flexible loops to be unresolved or make it difficult to unambiguously distinguish oxygen atoms and NH_2 groups. Additionally, identifying a target's binding site(s) is often made challenging by the flexibility of the receptor, whether it is amino acids in the binding pocket or entire loops occluding it. Furthermore, water molecules are crucial for ligand binding to a particular receptor but often not resolved by crystallography.

Techniques exist to address these problems in static and dynamic ways and all have implications for the accuracy of predictions regarding binding-site and receptor behaviour. Chemical probes can be used to identify binding 'hot spots' on a static 3D structure from X-ray crystallography (e.g. SiteMap [17], Fpocket [18]), multiple chemical probes applied to protein dynamics (e.g. MDMix [19]) normal mode analysis to determine a receptor's ensemble of most likely conformations ('micro states') as well as the dynamics between them [20] and, finally, using the coordinates of water in simulations transformed to distribution functions and discretized into a 3D grid from which thermodynamic quantities can be calculated, the thermodynamic profile of a binding pocket and implications for ligand binding characterised (e.g. Grid Inhomogeneous Solvation Theory [21])

Finally, accurately quantifying binding energies in an aqueous environment is largely out of reach even for the most accurate computational methods (e.g. Thermodynamic Integration). Protein-ligand binding affinity can be determined from the experimentally measured binding constant K_B

$$\Delta G = -RT \ln K_B = \Delta H - T \Delta S \quad (2.1)$$

Experimentally determined binding constants, K_B , are typically in the range of 10^{-2} - $10^{-12} \text{mol} \cdot \text{L}^{-1}$ and correspond to a Gibbs free energy of binding, ΔG , of between -10 and -70 kJ·mol⁻¹ in aqueous solution [22]. Two broad components addressed by the above relation are the ‘enthalpic’ and ‘entropic’ contributions to ligand binding, the former a function of making and breaking of electrostatic interactions between all three of protein, ligand and water with enthalpic contributions a function of the release of water to solvent and loss of conformational mobility of receptor and ligand.

Both enthalpic and entropic contributions often compensate between each other. Features that have been found to lead to more energetically favourable protein-ligand complexes include steric complementarity between protein and ligand, complementarity of surface properties (e.g. lipophilic parts of the ligand interact with lipophilic parts of the protein) and that the ligand adopts an energetically favourable conformation. However, as stated, binding involves a complex equilibrium between ensembles of solvated species varying in their conformation. In contrast, scoring functions that can be calculated quickly and efficiently use often crude approximations, use only one receptor-ligand geometry, neglect the unbound state of binding partners and often neglect cooperative effects of different interactions, usually assuming more of an additive relationship [23].

That said, at the high-throughput level SBVS inhabits, one can still use estimated energies for relative ranking of potential ligands as, although an accurate estimate is not possible, some correlations with experimentally-determined affinity are. Dozens of methods for estimating binding affinity of a protein-ligand complex have been developed since the early 90’s and a consequent taxonomy broadly groups them into five classes: (1) force field-based methods, (2) empirical scoring functions, (3) knowledge-based methods, (4) Quantum Mechanics (QM) scoring functions and (5) machine learning methods.

Force-field or ‘physics-based’ methods estimate affinity by complementing modelled estimates of gas-phase electrostatic interactions between protein and ligand and a pre-calculated cubic grid with solvation terms as estimated by continuum solvation models e.g. Generalised Born solvation model (GBSA). These methods fall under the general term of ‘molecular docking’ methods and have been implemented in UCSF’s DOCK [24], Autodock VINA [25] and OpenEye’s FRED [26]. Both receptor and ligand are rigid in these methods so they rely on input of multi-conformer ligands for fitting into the receptor’s binding pocket. Flexible

docking solutions exist, where an initial rigid fit is complemented with bond rotation and energy cut-offs [27] or incorporates both ligand and binding pocket flexibility via Genetic Algorithms [28]. However, these are often an order of magnitude slower than rigid methods and do not guarantee a significantly better result in all cases, such as with more diverse sets of molecules [29].

Empirical scoring functions estimate the binding free energy of a ligand by the summing of all individual free energies for interactions in a complex, where each function (f_i) of ligand (r_l) and protein (r_p) coordinates describing them is weighted by training sets of experimentally-determined binding constants.

$$\Delta G_{binding} \approx \sum_i \Delta G_i f_i(r_l, r_p) \quad (2.2)$$

Empirical scoring functions contain chemically-intuitive terms describing typical interactions (e.g. H-bonding, hydrophobic interactions, binding entropy) but also contain penalty functions for complexes that deviate from ideal values (e.g. atomic clashes, angle criteria).

Attempts have been made to improve scoring functions by weighting modalities of H-bonding [30] [31], ionic [32] and hydrophobic [33] interactions differently. This is as opposed to treating them as if they are identical or not significant in the overall binding, such as is the case with London or cation- π interactions. However, most methods tend to suffer from the same problem, that difficulty remains in predicting binding affinities of ligands that materially differ from those used in their training sets [34].

Knowledge-based scoring functions essentially sum all pairwise protein-ligand interactions where the potential between atom pairs i and j with distance r between them, $\omega_{ij}(r)$, is evaluated from the inverse Boltzmann law, the atomic densities are evaluated in pairwise fashion between the system state (ρ_{ij}) and a reference state and where interactions are assumed to be zero (ρ_{ij}^*).

$$\omega_{ij}(r) = -k_B T \ln[g_{ij}(r)] = -k_B T \ln \left[\frac{\rho_{ij}(r)}{\rho_{ij}^*} \right] \quad (2.3)$$

Using the assumption that a larger number of contacts between ligand and protein is indicative of attractive interactions, distributions of these contacts can be converted to atom-pair potentials, or, Potentials of Mean Force (PMF). Thus, in contrast to empirical scoring functions, that are derived from a training dataset generally obtained from the Protein Data

Bank (PDB), knowledge-based scoring functions make use of statistical analysis of structural information without the need for experimental binding data. Improvements have been observed by augmenting potentials with geometric [35] and solvent [36] data. Although fast, relatively straightforward to compute, and useful for re-ranking of docking results, no single scoring function consistently outperforms others when applied to diverse test sets with performance likely target-specific [37].

QM methods allow polarisation and charge-transfer effects of a region around the ligand to be modelled and incorporated into the energy terms of the scoring function [38]. More correctly under the QM/MM (Quantum Mechanics/Molecular Mechanics) approach, the ligand was treated with a relatively expensive linear-scaling semi-empirical Hamiltonian and the surrounding system with a cheaper classical Hamiltonian and the AMBER ff99SB force field. The effective Hamiltonian of the system was therefore a sum of the MM, QM and QM/MM Hamiltonians.

$$\hat{H}_{eff} = \hat{H}_{MM} + \hat{H}_{QM} + \hat{H}_{QM/MM} \quad (2.4)$$

Following energy minimisation steps, binding free energies were then calculated from the gas-phase binding free energy coupled with the solvation free energy.

Other approaches [39] integrating linear-scaling semi-empirical methods with the Linear Interaction Energy with Continuum Electrostatics (LIECE) model showed good agreement with experimentally-determined binding free energies of inhibitors against classes of proteins as diverse as an aspartic protease and a kinase.

The primary limitation for such methods is, however, throughput. Although certain optimisations drastically increased the speed of calculation (e.g. using energy minimisation instead of molecular dynamics), each molecule’s calculation took approximately an hour in 2008. Although computational speed has dramatically increased in this time, one could imagine that computations would remain of the order of tens of minutes per molecule, significantly slower than other scoring function approaches.

Finally, machine learning based scoring functions do not assume a functional form regarding binding affinity and structural features of the protein-ligand complex. Instead, they generate features directly from a training set of molecular data. Features generated from interatomic pairs of ligand and protein atoms are then selected via machine-learning algorithms (e.g. Random Forest, Support-Vector Machine), to build and validate a model to compute an outcome of interest, most commonly, binding affinity.

Machine learning scoring functions have been reported to outperform other classes of scoring function in predicting binding affinity of known protein-ligand complexes [40] even with less chemically discriminative models [41]. However caution must also be exercised as it has been reported that whilst producing excellent agreement with known complexes, machine learning methods have proven less suited to predict affinity for novel molecules to novel targets [42].

2.2 Fragment-based drug-discovery

One difficulty within the SBVS approach can lie within the molecular size of ligands. Screening with larger and more complex compounds can lead to a sort of molecular ‘over-fitting’ problem, where molecules are prioritised and selected for further development primarily because they are larger. This increases the probability of both false positives in computational studies and interactions with unintended targets in trials [43] [44].

In using one approach to solve this problem of larger false-positives, Fragment-Based Drug Design (FBDD), one starts small. Molecules of low molecular weight are screened against molecular targets. Although these molecules display lower overall affinity, the logic of FBDD follows that specific interactions from the fragment will result in a stable and predictable binding mode and that molecular extensions from this fragment will retain this binding mode in addition to the interactions provided by the extension. The approach also offers the proposition of an increase in the number of viable starting points in the binding pocket that should result in a greater diversity of molecules discovered and optimised to bind to a particular target. This potentially gives practitioners the opportunity to build desirable properties into a molecule from the beginning. Flowing from the FBDD philosophy, molecules should prioritise novelty, possess molecular properties useful in biological systems and, more recently, extensions should be relatively easy to synthesise.

A campaign to ensure chemical novelty for crystallographic fragment screens was conducted on molecules of less than 20 heavy atoms sourced from numerous commercial manufacturers [45]. This was born from the idea, later confirmed, that as fragments tend to be low-affinity binders, biophysical screening methods often miss molecules that are found to be hits via crystallography [46]. An initial library of > 200 000 fragments was reduced via several filtering steps, including the removal of reactive functional groups, toxic molecules and to ensure adherence to Lipinski’s ‘rule of 5’ [47]. From there, clustering steps to ensure chemical novelty pared the library down to 361 molecules. The library demonstrated a high

crystallographic hit-rate (19.7%) against an aspartic protease target and elucidated molecules with novel binding modes and chemotypes against this target [48].

Lower hit-rates but higher-affinity ligands were found in larger campaigns against multiple targets [49] but the trend is clear, that more libraries are being designed with chemical novelty in mind [50]. It does behove those looking for hits to be mindful of the biophysical screening methods used as it has been demonstrated that, whilst faster, assay methods (e.g. fluorescence assay, ElectroSpray Ionization-Mass Spectrometry) often result in vastly different hit rates with little overlap between them in identifying hits, in addition to not directly providing information about binding modes [51].

After finding fragments that bind, ligands can be extended, or, ‘elaborated’ in a number of ways. Successful examples of fragment linking or ‘target-guided synthesis’ have been reported, such as where multiple fragments binding to different areas of a receptor’s binding pocket are linked synthetically and then tested [52]. There are few successful campaigns in the literature, however, as designing a useful linker with appropriate flexibility has proven to be problematic. Too rigid and the original binding poses of each fragment are unlikely to be stable in the elaborated molecule. Too flexible and there will be an entropic penalty to pay that may abolish binding entirely.

Other approaches used, such as ‘tethering’, progress by synthesising disulfide-containing molecules that covalently bind to cysteines adjacent to the binding pocket. Tethering has been shown to effectively ‘probe’ a pocket and assist in finding an allosteric pocket, identify fragments suitable to interfere with Protein-Protein Interface (PPI) formation or to stabilise a complex that wouldn’t otherwise occur or form with much less stability [53]. The approach does demand specialised libraries of disulfide fragments [54] as well as the design and expression of multiple cysteine mutants for screening, limiting its utility.

In addition to structural information, one can be more concerned with synthetic feasibility of molecules from fragments known to bind to a target, as we were in chapter 3. The net effect is to increase the probability of hits, a philosophy under the broad methodology of *de novo* drug design. A library [55] of ‘synthetically tractable’ but drug-like molecules was also reported that resulted in a freely-available dataset of molecules derived from Chembridge [56] building blocks.

Beginning with relatively strict criteria for fragments ($\text{MW} \leq 250 \text{ Da}$, $n_{\text{rotatable bonds}} \leq 2$, $n_{\text{chiral centres}} \leq 1$) and reactions, a collection of 58 ‘robust’ reactions [57] was used to

exhaustively react ~8000 building blocks *in silico*. This generated ~21 million molecules with a focus on both chemical novelty and being drug-like but small enough to allow for further optimisation. Other approaches allow for customised libraries to be built or the reverse process, retro-synthetic deconstruction of a molecule or series of interest. This method also has, at its core, a focus on synthetic feasibility and from a small set of fragments, use of the approach suggested diverse extensions with up to a 40-fold improvement in affinity against the β_2 Adrenergic Receptor [58].

2.3 Molecular docking

The broad philosophy and techniques underpinning *in silico* ligand discovery efforts in chapter 3 and chapter 4 require intricate methods to estimate binding affinities and make full and proper use of structural information gained via experiment and developed molecular libraries. A common approach to facilitate the combination of both ligand and target structural information is molecular docking.

The principle inherent within molecular docking is to use the structure of a target protein to detect molecules active against it. In doing so, molecular docking takes advantage of the large and growing (>153 403 as at Jan 2019) amount of 3D protein structures made available via the Protein Data Bank (PDB) [59]. Broadly, this is achieved via the placement and scoring of molecular poses and then the evaluation of their fit in the binding pocket. Aside from marked differences in scoring functions, molecular docking programs are often characterised by two important facets; the degrees of freedom allowed within their algorithms, whether that be in the orientation or flexibility of the ligand, protein or both, and methods in how those degrees of freedom are searched. How solvation is handled and, for example, detection of atomic clashes also affects the accuracy of calculations. A meaningful taxonomy of docking software is beyond the scope of this thesis but two broad types used in chapter 3 and chapter 4 will be discussed.

2.3.1 Physics-based methods

The Solvation Energy for Exhaustive Docking (SEED) method used in chapter 3 places poses of fragment-sized molecules (≤ 250 Da) in a receptor binding pocket exhaustively in order to estimate the binding energy of each pose [60]. Pre-defined rules (e.g. dielectric value of solvent/solute, maximum angular deviation from ideal H-bond geometry) govern the generation and placement on binding pocket residues of vectors of uniform length on all

H-bonding groups in a direction of ideal H-bonding geometry.

A similar procedure occurs with polar groups on all pre-generated conformers of ligand molecules before matching of H-bond vectors is completed. Additional rotations around vector matches are performed to enhance sampling of conformational space in the binding pocket. Bad contacts are detected via the measurement of atomic van der Waals (vdW) distances and evaluation of binding energies to a pre-generated grid over the whole receptor.

After placement of a fragment pose, an evaluation of the total interaction energy is completed. Electrostatic energies in solution are calculated as the sum of all receptor and ligand desolvation energies and fragment-receptor interaction energies, evaluated atomically via Generalised Born approximation on atoms i and j of distance r in dielectric medium ϵ_p , or:

$$E_{ij}^{int} = \frac{q_i q_j}{\epsilon_p r_{ij}} - \frac{q_i q_j \tau}{R_{ij}^{GB}} \quad (2.5)$$

Where $\tau = (1/\epsilon_p) - (1/\epsilon_w)$ and where;

$$R_{ij}^{GB} = \sqrt{r_{ij}^2 + R_i^{eff} R_j^{eff} \exp\left(\frac{-r_{ij}^2}{4R_i^{eff} R_j^{eff}}\right)} \quad (2.6)$$

And where q_i and q_j are assigned partial charges and R_i^{eff} and R_j^{eff} are the effective Born radii of each atom [61].

Receptor desolvation is calculated via continuum electrostatics assumptions (e.g. uniform dielectric constant in solvent and solute regions). This negates the need for explicit treatment of solvent effects. Desolvation of the receptor is therefore the calculated energy difference between an uncharged fragment and a charged receptor in solution upon binding, or, the electric (Coulombic) displacement, \vec{D} , of a fragment of volume, V , on a 3D grid:

$$\Delta E = \frac{\tau}{8\pi} \sum_{k \in V_{frag}} \vec{D}(\vec{x}_k) \Delta V_k \quad (2.7)$$

Fragment desolvation is defined as the calculated energy difference between a charged fragment and uncharged receptor in solution upon binding, or atoms of partial charges q_i and q_j of distance r in dielectric medium ϵ_p ,

$$E = \sum_{e \in fragment} E_i^{self} + \sum_{i > j} \left(\frac{q_i q_j}{\epsilon_p r_{ij}} - \frac{q_i q_j \tau}{R_{ij}^{GB}} \right) \quad (2.8)$$

Where E_i^{self} is the interaction between the charge itself and solvent, non-zero for intra-molecular energies, and evaluated as:

$$E_i^{self} = \frac{q_i^2}{2R_i^{vdW}\epsilon_p} - \frac{q_i^2\tau}{2R_i^{eff}} \quad (2.9)$$

The difference in fragment energy is then evaluated simply as the energies of bound and unbound fragment in solution.

$$\Delta E = E^{docked} - E^{free} \quad (2.10)$$

An alternative approach by UCSF’s DOCK suite (chapter 3 and chapter 4) places ligands using a bipartite graph-matching algorithm to map subsets of atomic distances within a ligand to subsets of receptor-sphere internal distances [62]. Spheres are generated and placed into the binding pocket, reflecting groove and ridge complementarity of a pre-generated molecular surface [63] [64]. Complementary ligand orientations are then matched to the positions of these spheres (Figure 2.1). Separate electrostatic and vdW grids are calculated from the molecular surface and distance-dependent potentials between ligand and the nearest surface of the grids are calculated.

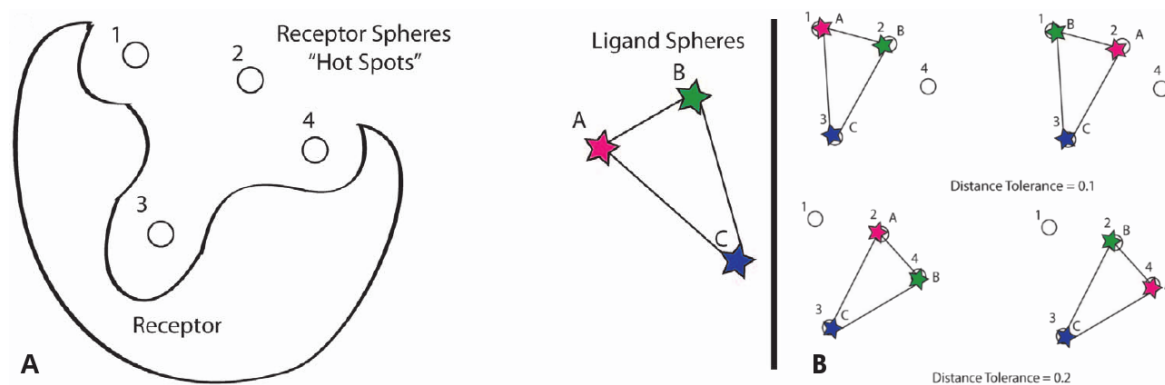


Figure 2.1: Depicted here is an example of the receptor:ligand sphere matching algorithm in UCSF DOCK. (A) A receptor with four spheres (empty circles) and ligand with three (coloured stars) is depicted prior to docking. (B) Four possible orientational matches of ligand spheres (coloured stars) at a given distance tolerance (in Å). Increasing the distance tolerance results in more orientations (i.e. with sphere 4) (from Coleman et al, 2013)

After ligand placement, a scoring function is applied to each atom in the pose comprising pre-computed partial charges interacting with the nearest grid points in pairwise fashion [65]. A vdW term is calculated from the AMBER united atom force field [66], an electrostatics term is calculated from a Poisson-Boltzmann calculation using electrostatics grids (QNIFFT [67]) and desolvation handled via Generalised Born approximation of the fractional desolvation

of all atoms in the pose [68]. The final score is then simply an additive relation combining each of these components.

$$E_{score} = E_{vdW} + E_{electrostatics} + E_{ligand,desolenergy} \quad (2.11)$$

Desolvation can also be estimated by applying the full transfer free energy, ΔG_i^{trans} , to each atom to calculate the overall free energy of desolvation of the ligand, ΔG_{desolv}^L .

$$\Delta G_{desolv}^L = \sum_i^{ligand} \Delta G_i^{trans} D_i \quad (2.12)$$

The fractional desolvation, D_i , over all atomic radii, a_i is then evaluated where dV is the volume of one grid element inside the binding pocket of the receptor.

$$D_i = \frac{a_i dV}{4\pi} \sum_k^{inr > a_i} \frac{1}{r_{ik}^4} \quad (2.13)$$

In contrast to SEED, DOCK assumes a lower dielectric constant inside the pocket than outside the pocket.

2.3.2 Shape-based methods

An approach from OpenEye is within the family of ‘exhaustive docking’. Given the hypothesis that the shape of the receptor is the most important factor in virtual screening, firstly, a negative image of the receptor’s binding pocket is generated by docking a set of drug-like molecular probes with only one conformer. The top-scoring poses of each probe are then ‘averaged’ to form a potential field that is converted to a density field by assuming a spherical Gaussian density distribution. The field created has higher potentials both at positions where ligand atoms make useful contacts with the receptor (e.g. polar H-bonds) and where ligand atoms are likely to occupy when molecules are large enough to stretch across multiple sub-pockets.

Unless pre-computed conformers are supplied, FRED [69] generates conformers then systematically enumerates all likely rotations and translations of each conformer within the binding pocket of the receptor. After generating an initial ensemble of poses and filtering all clear clashes using a negative image of the binding pocket, all remaining (tens-of-thousands to tens-of-millions) poses are scored and the top-100 are retained for the next step, a solid-body optimisation with 3^6 degrees of freedom. The top scoring pose is then kept as the final pose (Figure 2.2).

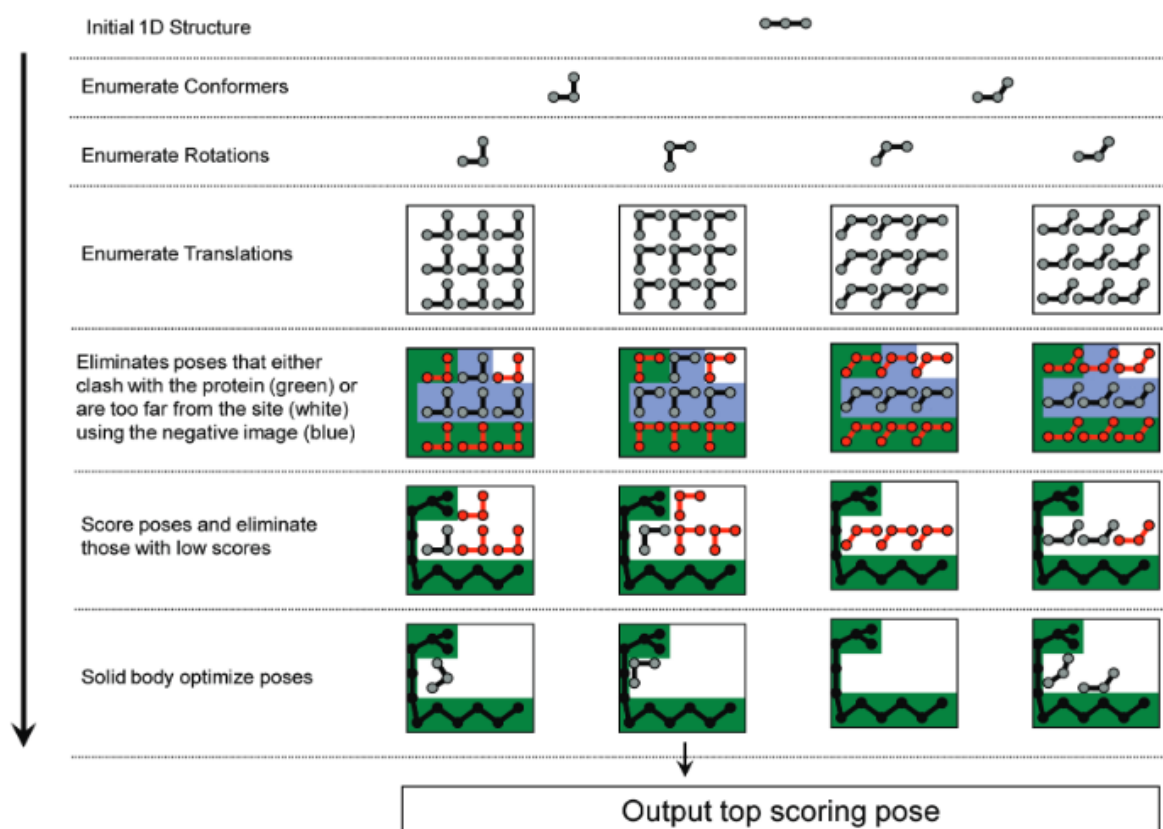


Figure 2.2: FRED docking process depicting the ligand scoring, optimisation and placement process. Conformers are exhaustively generated and progressive filtering steps applied as thousands of poses are scored and rejected based upon distances from the protein surface. Poses are then scored and a solid-body optimisation follows prior to presentation of the finalised pose position (from McGann, 2011).

FRED’s empirical scoring function, Chemgauss, models and smooths distance-based potentials for multiple types of interactions (e.g. H-bond, metal chelation) with additional facets modelled such as penalties for H-bonds with solvent broken by the docking process. A separate scoring function, Chemical Gaussian Overlay (CGO), models how well a pose overlays upon the structure of a known ligand, ensuring bias towards known binders in the docking process. This additional step is implemented in OpenEye’s HYBRID program and has been validated against a test set of 40 protein targets and multiple conformations of each protein accounts for binding site flexibility [26].

2.4 Molecular dynamics simulations

Whilst a valuable perspective, molecular docking where ligand and protein conformations are held rigid does not address dynamic protein and ligand behaviour at all. Similarly, flexible docking is limited to finding an energetic minimum for ligands and residues within range of them. The starting protein structures for these techniques are very often those obtained

from X-ray crystallography. The result is therefore more of an ‘averaged’ view of the system under study.

From the dynamics of a molecular system, making/breaking of hydrogen bonds, free energies, water interaction patterns, configurational changes (e.g. ring flips, puckering) and many other quantities can be estimated. The view of the energy landscape of protein folding has traditionally been that of the ‘funnel hypothesis’, that its native state corresponds to its free energy minimum in cellular solution conditions. However, at a larger scale, seminal papers [70] [71] have elucidated an understanding of protein systems dynamically sampling multiple conformational states, not necessarily all related to function. As the funnel hypothesis has given way to a more realistic description of the temporally-resolved motions of a system, more realistic treatments of chemically relevant quantities as they relate to time-resolved conformational movements of biomolecular assemblies has been the result [72].

Many techniques exist to alter the course of simulations and explore more conformational micro states, whether it be via biasing forces or guiding restraints. For example, in the generation of pathways through known endpoints of a system to gain information about allowed or disallowed states such as Targeted or Adaptive Steered Molecular Dynamics (T/SMD). Or, enhanced sampling of conformational states using techniques such as Umbrella Sampling, conformational flooding and Ensemble-Biased Metadynamics. For now, only unbiased (equilibrium) simulation methods as used in chapter 3 will be discussed.

2.4.1 Equilibrium simulation

The ingredients for a successful simulation are many and rooted in statistical mechanics, used to transpose information collected at the microscopic (atomic) level into macroscopic observables (e.g. potential energies, pressure, temperature). These can be used to study both thermodynamic properties of a system (ΔG) or kinetic (time-dependent) events.

By linking the macroscopic *thermodynamic state* of a system (Number of Particles \mathbf{N} , Temperature \mathbf{T} , Pressure \mathbf{P}) to its *microscopic state* (atomic positions \mathbf{q} and momenta \mathbf{p}), an *ensemble*, or a collection of micro states that all have the same constrained properties as the macroscopic system in multidimensional space (*phase space*) can be constructed. As a molecular dynamics simulation generates a sequence of points in phase space as a function of time, these points conform to different conformations of the system and belong to the same ensemble.

Depending upon the system under study, common ensembles include:

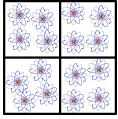
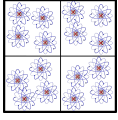
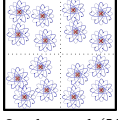
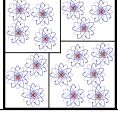
Ensemble	Constraints	Variables	Thermodynamic system
Microcanonical (NVE) 	Number of atoms \mathbf{N} Volume \mathbf{V} Internal energy \mathbf{E}	Energy of each particle ϵ_i	Isolated
Canonical (NVT) 	Number of atoms \mathbf{N} Volume \mathbf{V} Temperature \mathbf{T}	Internal energy \mathbf{E}_j	Closed
Grand canonical (μVT) 	Chemical potential μ Volume \mathbf{V} Temperature \mathbf{T}	Number of particles \mathbf{N}_i Internal energy \mathbf{E}_j	Open
Isothermal (NPT) 	Number of atoms \mathbf{N} Pressure \mathbf{P} Temperature \mathbf{T}	Volume \mathbf{V}_i Internal energy \mathbf{E}_j	-

Table 2.1: Comparison of various classical thermodynamic ensemble types, their properties and constraints in addition to the macroscopic thermodynamic system with which they may be compared (adapted from [73])

Broadly, if one allows a system to run indefinitely, then that system will eventually pass through all possible micro states. Thus, one aim of MD simulation is to generate enough representative conformations of a system such that a sufficient amount of phase space is sampled. This ensures that the link between macroscopic variables and microscopic states is established. An ensemble average is the mean of a quantity (e.g. temperature, pressure) that is a function of the micro state of a system. Ensemble averages, justified via agreement with experiment, are therefore calculated over a large number of replicas of the system.

The ensemble average where the observable quantity of interest, $A(p^N, r^N)$, is expressed as a function of all possible variables of the momenta p and positions r in the system, is given by:

$$\langle A \rangle = \iint dp^N dr^N A(p^N, r^N) \rho(p^N, r^N) \quad (2.14)$$

The probability density of the ensemble with Hamiltonian H , Temperature T , Boltzmann Constant k_B and partition function Q is then:

$$\rho(p^N, r^N) = \frac{1}{Q} \exp\left[-\frac{H(p^N, r^N)}{k_B T}\right] \quad (2.15)$$

Where the partition function can be evaluated as:

$$Q = \iint dp^N dr^N \exp\left[-\frac{H(p^N, r^N)}{k_B T}\right] \quad (2.16)$$

This integral relies upon the availability of all possible micro states and a MD simulation passing through all of these states. Instead, MD simulations can be used to calculate a time-related ensemble average. So, a simulation of time t , number of steps M and instantaneous value of $A(p^N, r^N)$ can be used to calculate the time-averaged value of A .

$$\langle A \rangle_{time} = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A(p^N(t), r^N(t)) dt \approx \frac{1}{M} \sum_{t=1}^M A(p^N, r^N) \quad (2.17)$$

This leads to the insight of the Ergodic Hypothesis that the time-average of A equals the ensemble average:

$$\langle A \rangle_{ensemble} = \langle A \rangle_{time} \quad (2.18)$$

With the assumption that the time average of a sufficiently long simulation is equal to the ensemble average of thermodynamic quantities, we now have a basis for conducting MD simulation to obtain said quantities.

In practice, the act of conducting a MD simulation is based upon Newton's Second Law of Motion where the total force acting upon atoms are related to their mass m and acceleration a .

$$F_i = m_i a_i \quad (2.19)$$

Integration of the equations of motion then allow us to determine positions, accelerations and velocities of atoms as they vary over time (*trajectory*). So for atoms of velocity v and position x at constant acceleration a :

$$F = m \cdot a = m \cdot \frac{dv}{dt} = m \cdot \frac{d^2x}{dt^2} \quad (2.20)$$

After integration, expressions for time-dependent atomic velocities and positions from initial velocities v_0 and positions x_0 can be stated as:

$$v = at + v_0 \quad (2.21)$$

And:

$$x = v \cdot t + x_0 \quad (2.22)$$

Combined, an expression is yielded to track changes in positions of atoms over time from initial positions and velocities:

$$x = \frac{1}{2} \cdot a \cdot t^2 + v_0 \cdot t + x_0 \quad (2.23)$$

With acceleration given as a derivative of potential energy E with respect to position r :

$$a = \frac{1}{M} \frac{dE}{dr} \quad (2.24)$$

Initial positions are usually chosen from experimentally determined structures (e.g. X-ray or NMR spectroscopy). Initial velocities v_i are usually selected randomly from a Maxwell-Boltzmann or Gaussian distribution of temperature T that yields the probability that an atom i has a velocity v_x in the direction of x .

$$p(v_{ix}) = \left(\frac{m_i}{2\pi k_B T} \right)^{\frac{1}{2}} \exp \left[-\frac{1}{2} \frac{m_i v_{ix}^2}{k_B T} \right] \quad (2.25)$$

Where the temperature can be calculated for a system of number of atoms N .

$$T = \frac{1}{3N} \sum_{i=1}^N \frac{|p_i|^2}{m_i} \quad (2.26)$$

As the potential energy, a function of atomic positions ($3N$), must be calculated analytically, several popular algorithms exist. The algorithm used most commonly by the Ambertools simulation software is the Leap-frog algorithm. Velocities are first calculated at time $t + \frac{1}{2}\delta t$ which are then used to calculate positions r at time $t + \delta t$

$$v(t) = \frac{1}{2} \left[v \left(t - \frac{1}{2}\delta t \right) + v \left(t + \frac{1}{2}\delta t \right) \right] \quad (2.27)$$

From the integration of the equations of motion, an estimate of the energy of a system may be obtained via a potential energy function, or, force field. There are many force fields that introduce approximations and impose limitations on their use with certain systems.

For example, specific force fields govern the motion and boundary conditions of everything from small molecules to large protein systems [74, 75] where secondary structure inherent in protein systems means dihedral corrections and rotation barriers for certain residues under certain circumstances must be applied. The parametrisation of these more complex force fields coupled with the number of variants are beyond general description here.

For now, as a general example of a well-developed general purpose force field, the General Amber Force Field (GAFF [76]) shall be introduced. A simple harmonic function form calculates the bond and non-bonded energy of a given atom pair (i, j) in media of dielectric constant ϵ . Partial charges (q_i, q_j) , bond angles θ separated by distance R are augmented by empirically-determined force parameters K_r , K_θ , V_n and equilibrium structural parameters r_{eq} and q_{eq} .

$$E_{pair} = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right] \quad (2.28)$$

All of the above are implemented in CPU-bound and GPU-accelerated codes in the Amber-tools [77] suite of programs used in chapter 3 and from here, the first stages of a simulation can begin.

First, an energy-minimised starting structure is solvated via the overlay of water molecules to completely surround the solute.

A *heating* stage follows where initial velocities are assigned to atoms in the system and water molecules allowed to move. The system is heated from low temperature and the movements of the solute restrained to minimise large forces that may lead to instability. As the system is heated and the system stabilises, restraints on the solute are loosened and ultimately lifted entirely as the system reaches the target temperature.

The purpose of the next stage, *equilibration*, is to allow the system to stabilise with respect to quantities such as temperature, pressure and energy over time. It is run for as long as it takes for this to occur, typically a few nanoseconds.

The final stage, *production*, is run for as long as is desired to observe the chemical phenomena in question. In this phase, thermodynamic parameters are calculated so the length of the simulation is dependent upon the ensemble as identified earlier and whether adequate sampling of conformations have occurred to ensure the accuracy of these quantities.

2.5 Energy minimisation

Primarily, the goal of energy minimisation in chemistry is to determine a set of coordinates representing a molecular conformation such that the potential energy of the system is at a

minimum on the potential energy surface describing the molecule. Several algorithms are implemented in chemistry software packages, generally working in Cartesian space, with varying convergence criteria, complexity and tolerances for how far a molecule really is from a local minimum, or, stationary point. Differences in a given system govern which method is most appropriate.

For now, only two first-order minimisation methods, *Steepest Descent* and *Conjugate Gradient*, both implemented in the Chemistry at HARvard Macromolecular Mechanics (CHARMM [78]) software package, will be discussed. Also discussed will be an approach combining a rapid method of estimating ligand conformational entropy with a second-order optimisation, implemented in the OpenEye software SZYBKI. All of these approaches were used to find energetic minima of proteins and protein-ligand interactions in chapter 3 and chapter 4.

2.5.1 Steepest descent minimisation

First introduced by Cauchy [79], the aim of the steepest descent, or, gradient descent method is to find the minimum of a function $f(x), x \in R^N$ and that $f : R^N \rightarrow R$ of gradient $g_k = g(x_k) = \nabla f(x_k)$ [80]. The minimum of a given quadratic function is found by computing steps along a given search direction d_k .

$$x_{k+1} = x_k + \alpha_k d_k, k = 0, 1, \dots, n \quad (2.29)$$

Where the step length α_k is given such that:

$$\alpha_k = \operatorname{argmin} f(x_k) + \alpha d_k \quad (2.30)$$

For the steepest descent method, the search direction is given as where the slope is negative, or, $d_k = -\nabla f(x_k)$. The algorithm therefore proceeds until a given convergence tolerance is reached or variations fall below a given tolerance.

Algorithm 1 Steepest descent algorithm

Input: Given an initial x_0 , $d_0 = -g_0$ and convergence tolerance tol

```

1: for  $k = 0$  to  $maxiter$  do
2:   Set  $\alpha_k = argmin_{\alpha} \phi(\alpha) = f(x_k - \alpha g_k)$ 
3:    $x_{k+1} = x_k - \alpha_k g_k$ 
4:   Compute  $g_{k+1} = \nabla f(x_{k+1})$ 
5:   if then  $\|g_{k+1}\| \leq tol$ 
6:     converged
7:   end if
8: end for

```

The CHARMM implementation of steepest descent adjusts the step size as the calculation approaches convergence. For example, if a step results in a lower energy, the step size is increased by 20% whereas if the energy rises, the step size is halved until convergence is reached.

2.5.2 Conjugate gradient minimisation

From a similar starting point to the steepest descent, the conjugate gradient method [81] uses the property of vector conjugacy of a set of nonzero vectors $\{p_0, p_1, \dots, p_{n-1}\}$ to a Symmetric Positive Definite (SPD) matrix A (*A-conjugacy*).

$$p_i^T A p_j = 0 \quad (2.31)$$

This set of A-conjugate vectors acts as a basis for the whole of real number space \mathbb{R}^N and can be used to express the difference between the exact solution x^* and the first guess x_0 as a linear combination of these conjugate vectors.

$$x_* - x_0 = \sigma_0 p_0 + \sigma_1 p_1 + \dots + \sigma_{n-1} p_{n-1} \quad (2.32)$$

By using this property of conjugacy, the coefficients σ_k are equivalent to the step size α_k needed to minimise a quadratic function along $x_k + \alpha_k p_k$ so the exact solution becomes a linear combination of the initial guess coupled with step size and the conjugate vectors.

$$x_* = x_0 + \alpha_0 p_0 + \alpha_1 p_1 + \dots + \alpha_{n-1} p_{n-1} \quad (2.33)$$

This only demonstrates that, in theory, a set of A-conjugate vectors exists. To each successive step, each new direction p_k is chosen as a linear combination of the negative residual $-r_k$

(equal to the steepest descent, or, negative gradient direction) and *only* the previous search vector p_{k-1} .

$$p_k = -r_k + \beta_k p_{k-1} \quad (2.34)$$

Where β_k is found by imposing a conjugacy condition, $p_{k-1}^T A p_k$, upon the current step by the previous step.

$$\beta_k = \frac{r_k^T A p_{k-1}}{p_{k-1}^T A p_{k-1}} = \frac{r_k^T r_k}{r_{k-1}^T r_{k-1}} \quad (2.35)$$

The algorithm therefore becomes an exercise in using the previous step as a guide toward the nearest minimum. The step length is calculated (step 2) then the solution is updated with that step (step 3). The residual is updated (step 4) and then, finally, the new search direction is obtained in the final two steps (5 and 6).

Algorithm 2 Conjugate gradient algorithm

Input: $r_0 = Ax_0 - b$, $p_0 = -r_0$

1: **for** $k = 0, 1, 2, \dots$ until convergence **do**

2: $\alpha_k = \frac{r_k^T r_k}{p_k^T A p_k}$

3: $x_{k+1} = x_k + \alpha_k p_k$

4: $r_{k+1} = r_k + \alpha_k A p_k$

5: $\beta_k = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$

6: $p_{k+1} = -r_{k+1} + \beta_k p_k$

7: **end for**

As the step size is determined by the algorithm itself, in CHARMM it is implemented as an automatic step. In terms of minimising the number of steps taken to reach convergence, the conjugate gradient method will significantly outperform steepest descent. However, with poor conformations, it will generate more integer overflows. Hence, it is common practice to apply both algorithms to a given structure, beginning with steepest descent minimisation followed by conjugate gradient minimisation as it nears the energetic minimum.

2.5.3 Minimisation and ligand entropy

Estimation of ligand entropy is essential to accurate estimation of the binding free energy in a protein-ligand environment. However, accurate condensed-phase calculations of the conformational entropy are computationally expensive, usually determined by MD simulation assuming a trajectory is run for long enough. As accurate prediction of molecular entropy

relies upon a given molecule being at its energetic minimum, a fast but relatively accurate treatment of ligand entropy in a protein-ligand environment [82] is necessary in drug-design research where the number of ligands being evaluated is often large.

Ligand entropy is estimated as the sum of configurational entropy S_c and solvation entropy ΔS_s

$$S = S_c + \Delta S_s \quad (2.36)$$

Configurational entropy

Configurational entropy is estimated as a relation to a partition function q with Boltzmann constant k , number of ligands N , and absolute temperature T .

$$S_c = kN \left[1 + \ln \left(\frac{q}{N} \right) + \frac{T}{q} \frac{\delta q}{\delta T} \right] \quad (2.37)$$

The conformationally-dependent partition function q of conformer i with internal energy ϵ_i , translational, rotational and vibrational partition functions (q_t , $q_i r$ and $q_i v$ respectively) and number of conformers n_c can therefore be evaluated.

$$q = q_t \sum_{i=1}^{n_c} e^{-\frac{\epsilon_i}{kT}} q_i r q_i v \quad (2.38)$$

Introduction of a simplifying assumption, that a ligand in a protein-ligand complex has three translational and three rotational degrees of freedom, mean q can be evaluated as either q_v for a single binding mode or a function of all q_{iv} 's for the number of binding modes n_p .

$$q = \sum_{i=1}^{n_p} \exp \left(-\frac{\epsilon_i}{kT} \right) q_i v \quad (2.39)$$

Vibrational frequencies underpinning the evaluation of q_{iv} are evaluated from normal-mode analysis of the relation between mass-weighted second-derivative Hessian matrices H^m , the true Hessian \mathbf{H} and atomic mass matrix \mathbf{M} .

$$H^m = M^{-\frac{1}{2}} H M^{-\frac{1}{2}} \quad (2.40)$$

Where eigenvalues λ_i of H^m can be used to determine harmonic wave numbers \tilde{V}_i .

$$\tilde{V}_i = \frac{1}{2\pi c} \sqrt{\lambda_i} \quad (2.41)$$

The second-order Broyden-Fletcher-Goldfarb-Shanno (BFGS [83]) minimiser is used to construct a diagonalised matrix H_{i+1}^{-1} after each step i from atomic position vector x_i to predict the positions of atoms in the next step x_{i+1} , where $f(x_i)$ is a function governing where the desired minimum is.

$$x_{i+1} = x_i + H_{i+1}^{-1} \nabla f(x_i) \quad (2.42)$$

In words, BFGS optimisation, especially limited-memory (L-BFGS) as implemented in Szybki, is a way of finding a (local) minimum of an objective function, in this case $f(x_i)$, making use of the objective function values and its gradient ∇ from a starting point of an initial diagonalised Hessian. So, with every iteration, the approximate Hessian approaches the exact Hessian. Upon reaching a good quality Hessian, the optimisation is complete and a minimum established by calculation of normal frequencies.

Primarily, there are two advantages to second-order minimisation methods using the Hessian and gradient over first-order methods using the gradient. Firstly, speed of convergence. Algorithms such as Steepest Descent tend to exhibit linear convergence [84] whereas L-BFGS has superlinear, and therefore much faster, convergence [85]. Secondly, L-BFGS tends to outperform both Steepest Descent and Conjugate Gradient on low-dimensional problems ($< 10\,000$ parameters), especially if the starting point of the optimisation is far away from a local minimum [86].

Solvation entropy

Solvation entropy ΔS_s is estimated from an additive relation between electrostatic ($\Delta S_{s,elec}$) and hydrophobic ($\Delta S_{s,hyd}$) components.

$$\Delta S_s = \Delta S_{s,elec} + \Delta S_{s,hyd} \quad (2.43)$$

The electrostatics component is further broken down into bulk effects and polar solute-solvent effects of dielectric constants ϵ_{solv} for each component and temperature T for the second component.

$$\Delta S_{s,elec,bulk} = - \left(\frac{\delta \Delta G_s}{\delta \epsilon_{solv}} \right) \left(\frac{\delta \epsilon_{solv}}{\delta T} \right) \quad (2.44)$$

The second term is evaluated at a constant of 28 J.molK^{-1} based upon various gas-phase calculations of molecules with proton donors or acceptors interacting with water molecules.

Accurate treatment of specific hydrophilic solute-solvent clusters was determined to be beyond the scope of the method, making Szybki computationally too expensive for the purposes of drug design research.

Finally, the hydrophobic term is evaluated in approximate fashion from the hydrophobic solvation free energy $\Delta G_{s,hyd}$ in Temperature T .

$$\Delta S_{s,hyd} = - \left(\frac{\delta \Delta G_{s,hyd}}{\delta T} \right) \quad (2.45)$$

Where $\Delta G_{s,hyd}$ is an additive relation of the free energies of cavitation ΔG_{cav} , solute-solvent vdW ΔG_{vdW} and induction (permanent dipole-induced dipole) ΔG_{ind} .

$$\Delta G_{s,hyd} = \Delta G_{cav} + \Delta G_{vdW} + \Delta G_{ind} \quad (2.46)$$

All three free energies are estimated from approximate methods to allow for the packing density solvent around the ligand (e.g. Scaled Particle Theory, or, SPT) as well as analytically-determined relations that calculate solvent and ligand polarisabilities and dipole moments.

2.6 Ligand-based techniques

The use of ligand-based techniques in drug design has previously been a very fertile area of research [87]. The broad philosophy is used to develop drug candidates in the absence of structural information describing the target and is predicated on the assumption that structurally similar molecules will have similar biological activity [88]. An outgrowth of this has been statistical and analytical tools to elucidate the structural parameters and descriptors that link a given molecule or molecular class to biological activity.

A popular family of techniques to address this are Quantitative Structure-Activity Relationships (QSAR [89] [90]). The broad method involves *in silico* modelling of candidate ligands that are energy minimised. Molecular descriptors for ligands are generated and are often structural, physicochemical and/or based on calculations from knowledge-based, molecular mechanics or quantum mechanical tools. A mathematical relation is developed to explain variation in a dependent variable, that being biological activity against the target. The model is then tested, refined and validated via an array of techniques more usually associated with linear regression analysis.

Other approaches, such as the 3D Pharmacophore model [91], focus primarily on the ligand and structure but integrate some structural information. To that end, a 3D pharmacophore

model is an essential set of interactions aligned in three-dimensional space of small-molecule ligands with a protein structure. These interactions encompass intuitive chemical descriptors such as hydrogen-bond donors and acceptors but also integrate harder to elucidate interactions such as charge-transfer interactions and electrostatic interactions with e.g. Fluorine atoms in a ligand. Variants on this, such as the Conformationally-Sampled Pharmacophore approach (CSP [92]) integrate molecular dynamics simulation as an initial step to increase the probability of finding the receptor-bound form of the ligand.

2.6.1 Similarity and fingerprinting

Another outgrowth of ligand-based methods is in molecular 2D similarity, or, ‘fingerprints’. The development of a fingerprint generally involves the encoding of 2D chemical structure information (e.g. number of rotatable bonds, number of atoms) in a bit-string. This is then compared to others via Tanimoto coefficient as a measure to determine the similarity between comparison molecules [93].

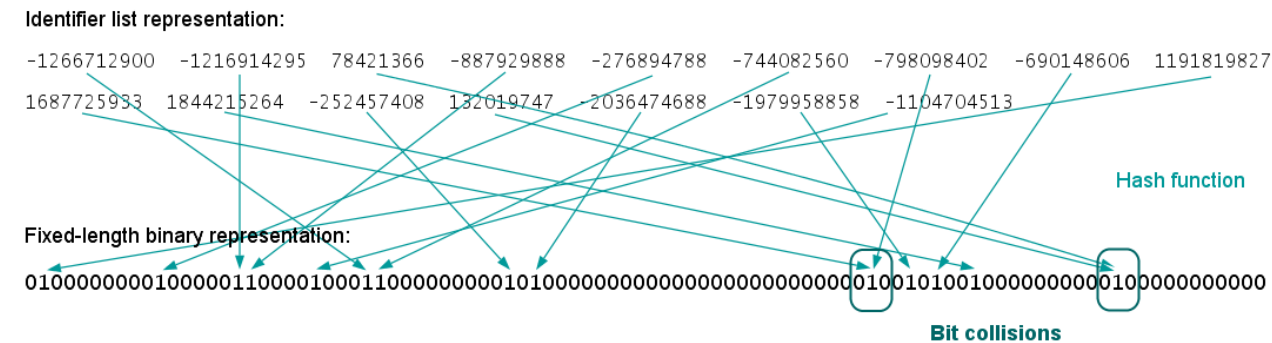
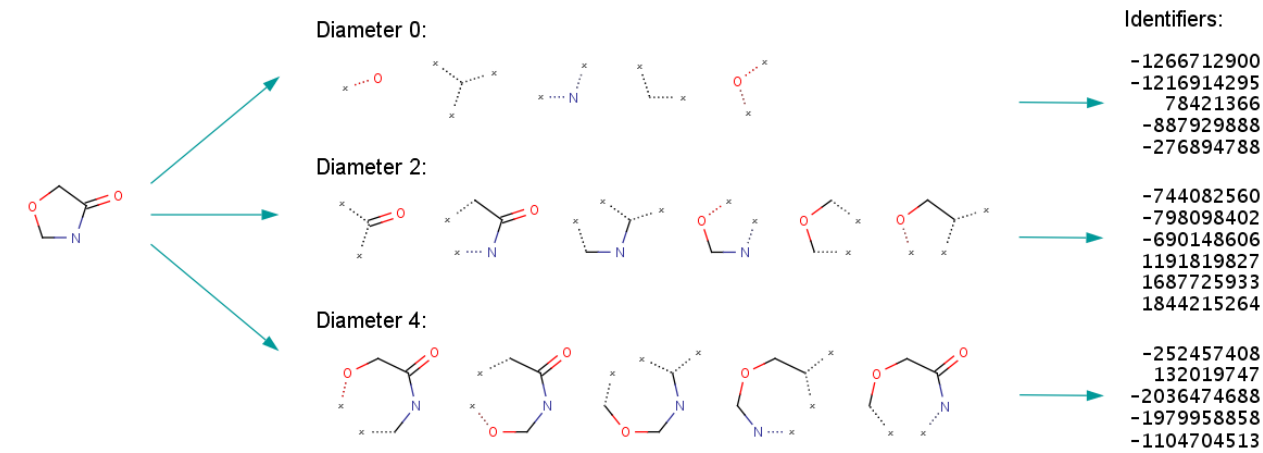
The Tanimoto coefficient τ is evaluated where N_A , N_B and $N_{A\&B}$ is the number of features in the fingerprints of query molecule A and B individually and bits that are common, respectively.

$$\tau = \frac{N_{A\&B}}{N_A + N_B - N_{A\&B}} \quad (2.47)$$

An attempt to integrate more molecular features into molecular fingerprints resulted in the Extended-Connectivity Finger Print (ECFP [94]) method, a variant of the Morgan algorithm [95] and explicitly designed to capture features relevant to activity.

The ECFP algorithm firstly assigns integers to the atoms in the query molecule prior to an iterative updating stage where an atom’s identifier is updated to reflect the identifiers of their neighbours. This includes basic chemical information (e.g. attached bond types, whether part of a ring, atomic mass) followed by a duplicate removal stage (Figure 2.3). The final stage is ‘folding’, a method to increase the information density stored in a fingerprint. With fixed-length comparisons, strings that could contain most conceivable chemical moieties, fingerprints for small molecules will contain many 0’s and therefore be ‘sparse’ whereas for very large molecules, the bit string will contain many 1’s and therefore be ‘rich’. This was not considered an efficient way to store molecular representations.

The folding process begins with a fixed fingerprint size that is very large to accurately represent any molecule expected to be encoded. The fingerprint is then ‘folded’ by dividing



Chapter 3

Novel ligands for the kinase

Proviral Integration site for MuLV

(Pim-1)

In conjunction with my supervisor, Prof. Dr. Peter Kolb, I was responsible for the overall research planning, strategy and execution of all projects. This is with the exception of Project 4 where Prof. Dr. Diederich noted the significance of the X-ray crystal structure that began the project. I was responsible for completing all computational work with the exception of the docking of aniline-triazoles in Project 2 and library collation for Project 4, both completed by Dr. Florent Chevillard. Chemical synthesis of all compounds was carried out by Mr. Lukas Heyder from our collaborators, AG Diederich. I was responsible for assay testing of compounds with help from Mr. Matthias Oebbeke with exception of those purchased for Project 3 that were also tested by Eurofins Inc. All Thermal Shift Assays and X-ray crystallography were carried out by Dr. Christof Siefker.

3.1 Introduction

3.1.1 The kinase Pim-1

The Proviral Integration site for MuLV (PIM) kinases are a family of serine/threonine kinases in the CalModulin-Dependent Kinase (CAMK) family.

PIM genes were first identified in the 1980s as potent oncogenes associated with murine leukaemia virus (MuLV)-induced lymphomas in mice [97] and comprise a family of three genes: Pim-1, Pim-2 and Pim-3. These genes have differing tissue expression profiles [98] and a sequence homology of >60%. This leads to compensatory activation [99] in the absence of one or more of the proteins expressed by PIM genes but particularly Pim-1 [100].

Pim-1 interacts with and phosphorylates a variety of cellular targets and performs a variety of roles involved in cell cycle progression and death. Pim-1 inhibits apoptosis via interactions with regulator proteins bcl-2 and Gfi-1 [101] but has arguably a more significant role in a pathway involving the Janus Kinases and Signal Transducer and Activator of Transcription proteins (JAK-STAT pathway [102]).

The JAK-STAT pathway (Figure 3.1) is involved in cellular processes leading to DNA transcription such as cell survival, proliferation and differentiation [103]. Pim-1 has cell cycle substrates necessary for mitosis (NuMA [104]), cyclin-dependent kinase inhibition to regulate DNA synthesis (p21^{Cip1}) [105], inhibitors (p27^{Kip1} [106]) and activators (Cdc25A [107] and Cdc25c [108]) of cell division. Pim-1 also has a role inhibiting the formation of a heterodimer between Bcl-2 and Bcl-2-associated death promoter (BAD) via phosphorylation of S112 and S136 in the latter, leading to inhibition of Bak/Bak-triggered apoptosis [109].

Likely the most potent collaboration, however, is between Pim-1 and a transcription factor, c-Myc [110]. c-Myc regulates many important genes and processes (e.g. telomerase activity, differentiation) and its dysregulation, most commonly via amplification, is associated with most cancers [111]. c-Myc typically has a very short half-life of 10 to 20 minutes prior to its dephosphorylation and ubiquitination as part of one or more Raf pathways [112]. However, in one of the most compelling collaborations described between any combination of oncogenes, it has been established that even mild Pim-1 over-expression and increased levels of c-Myc are associated with highly malignant monoclonal tumours [113], despite being relatively weak oncogenes individually. Pim-1 inhibits c-Myc degradation via increased phosphorylation at two key sites in c-Myc (S62 and S329), found via protein stability assays to dramatically increase the half-life of c-Myc to >90 minutes [114].

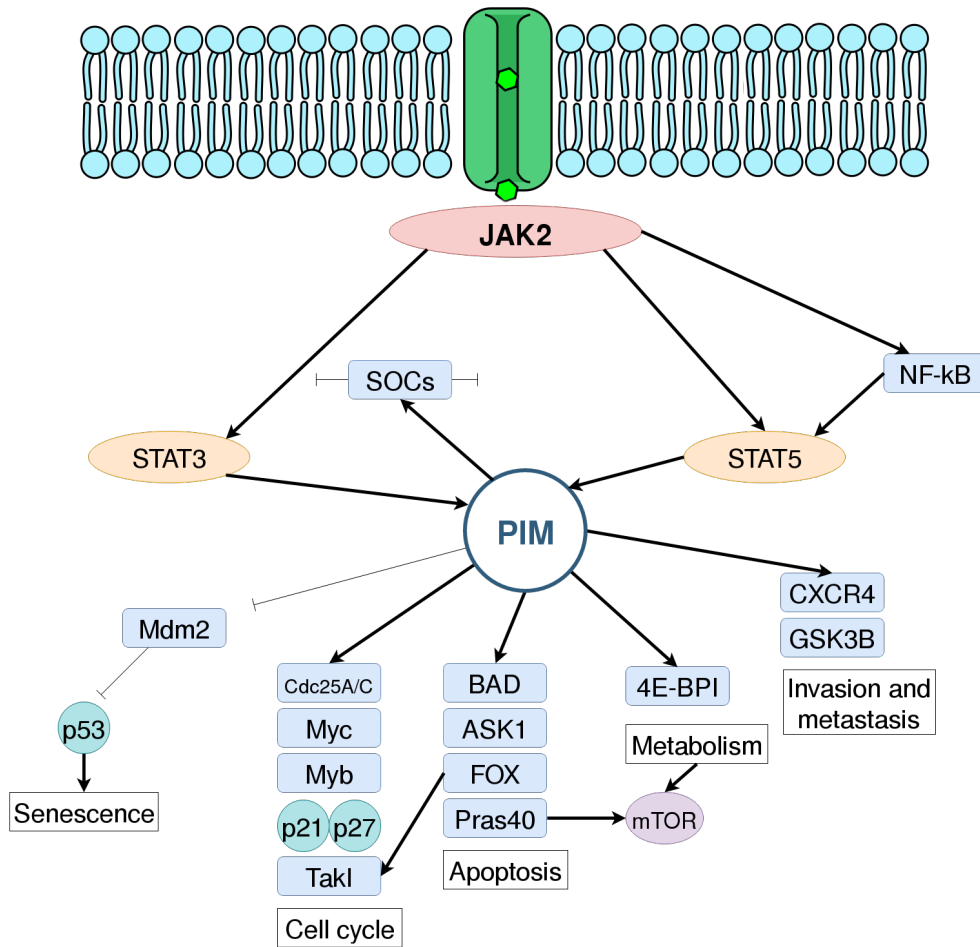


Figure 3.1: Pim-1 in the JAK/STAT pathway. Upon activation by cytokine binding to cellular receptors (green cylinder), JAK2 (pink ovoid sphere) autophosphorylates its cytoplasmic receptor domain, creating recruitment sites for STATS (yellow ovoid spheres). STATS 3 and 5 directly bind to the Pim-1 promoter regions and up-regulate its gene expression. Expression of all three PIM proteins are also induced by activation of transcription factor complexes such as NF-kB. Additionally, Pim-1 is also able to negatively regulate the JAK/STAT pathway via binding to SOCs, a group of negative regulators. PIM in the JAK2 pathway subsequently phosphorylates, and consequently activates or inactivates, a number of enzymes and transcription factors (light blue ovoid spheres) and kinase inhibitors (aqua spheres) in many cellular pathways (adapted from [102]).

Consequently, Pim-1 kinases are overexpressed, not just in a variety of haematopoietic malignancies but also solid tumours. This often leads to cancers in the prostate [115], lungs, gastric system, pancreas and squamous cell carcinomas of the upper-body [116].

As levels of Pim-1 are effectively rate-limiting for tumour progression in certain lymphomas, this makes it an attractive target for drug therapies against these diseases. Some unique structural factors that make Pim-1 even more apposite a drug target are that it is constitutively active, requiring no further post-translational modification once properly folded. Pim-1's cellular activity is also regulated by its own synthesis and degradation, hence, expression and activity are predictable [117].

Structurally, Pim-1 possesses the usual kinase features such as a characteristic bi-lobular fold comprising the N-terminus and C-terminus is augmented by a catalytic domain comprised of an activation loop (DFG loop), a ‘hinge’ region generally occupied by Adenosine Tri-Phosphate (ATP) and a flexible two-stranded β -sheet (Glycine-rich or P-Loop). Within the binding pocket are two residues key for kinase inhibitor binding. The adenine moiety of ATP interacts with E121 in the Hinge so inhibitors targeting this residue are known as ‘ATP-mimetic’ whereas the opposite side of the pocket contains a catalytic [118] Lysine (K67) that interacts with inhibitors in ‘ATP-competitive’ fashion (Figure 3.2).

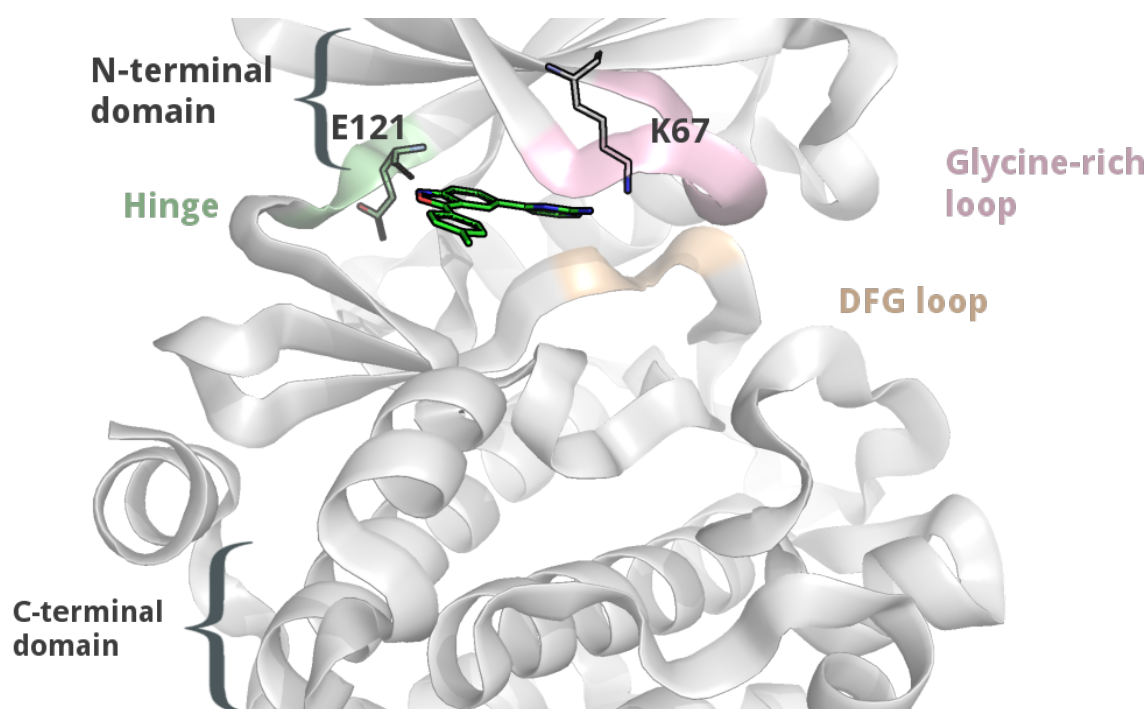


Figure 3.2: The basic structure of the kinase Pim-1. Depicted here are the N-terminus and C-terminus (grey cartoon) signifying the respective beginning and end of the protein sequence, a flexible Glycine-rich P-Loop (pink cartoon), the DFG loop (straw cartoon) that adopts alternative conformations delineating activation or inactivation, a conserved Lysine (K67, stick representation) that plays a catalytic role in phosphate transfer from bound ATP to Pim-1’s co-substrate, a hinge region (green cartoon) that directly binds the adenine moiety of ATP via E121 (stick representation) and a bound ligand (PDB:3BGP, green carbons, stick representation) displaying a binding mode usually associated with ‘ATP-mimetic’ inhibition.

As with other kinases, Pim-1 has several regions where non-conserved residues are found. This arises because ATP does not occupy the entire volume of the binding pocket so regions around the segments of a bound ATP molecule have been named accordingly. Two of the more important areas are the Hydrophobic Pocket, situated behind the adenine moiety of ATP and the Specificity Surface, situated at the opening of the binding pocket [119]. Both areas are believed to be important for kinase selectivity (Figure 3.3).

PIM inhibition has been intensely studied for over a decade and highly diverse classes of

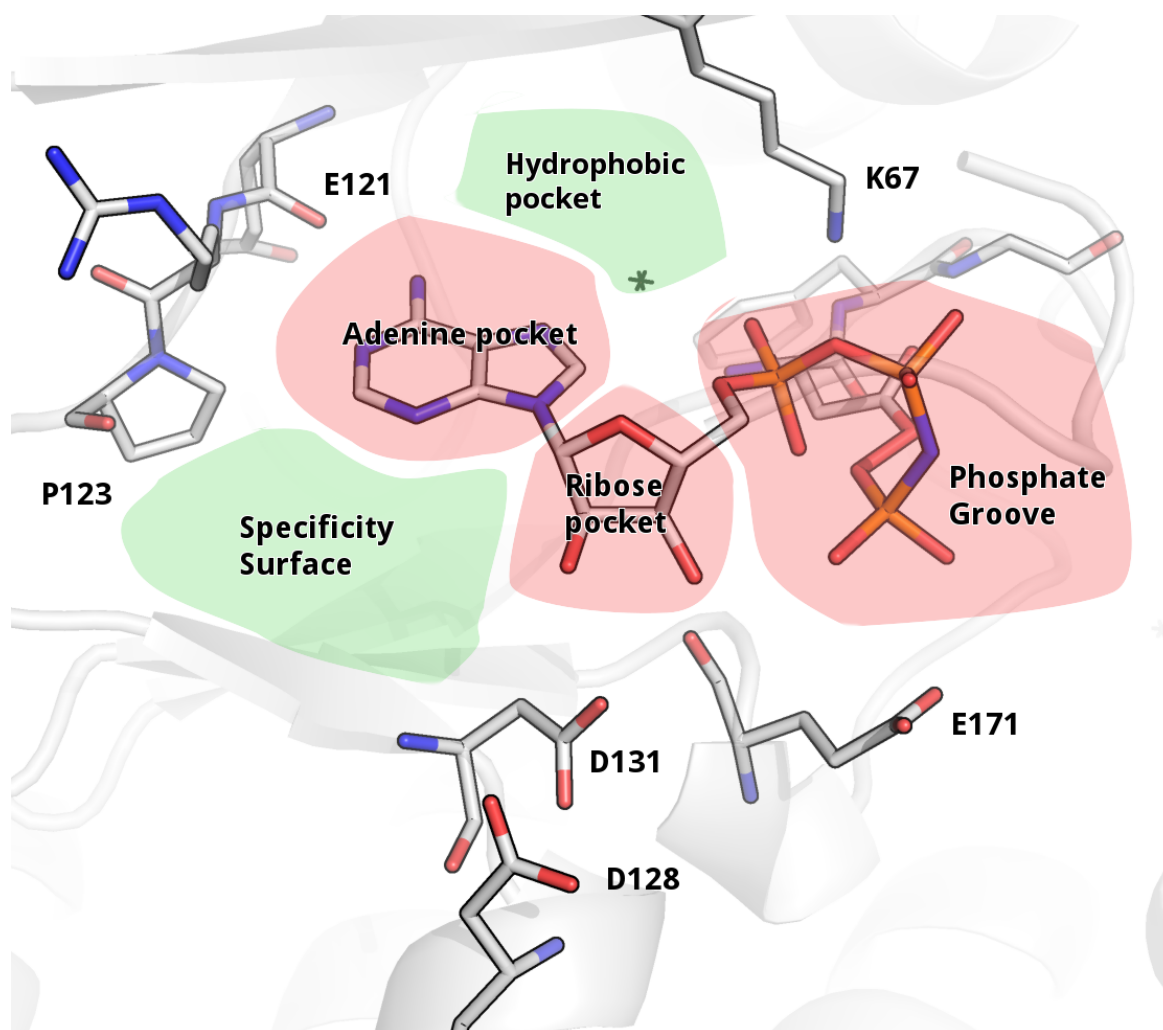


Figure 3.3: The Pim-1 binding pocket showing the protein in complex with AMPPNP (PDB:1YXT) and non-conserved regions categorised by ATP binding mode, green areas where apolar binding occurs and red where polar interactions occur. Key residues (e.g. E121, D128) are depicted by light-grey stick representations.

inhibitors have been published [120] displaying a few important structural features. Generally, ligands have possessed a rigid core to fill the ATP binding site, interacting with apolar residues (e.g. I104, F49). Polar interactions have generally been formed with residues in a hydrogen-bonding network comprised of a catalytic lysine (K67) and residues in the DFG loop (D186).

So-called ‘hinge binders’ have been situated in the adenine pocket. However, Pim-1 is uniquely apolar when compared to other kinases. Hence, where there are at least two polar bonds available in the hinge in other kinases, the position and geometry of a Proline (P123) in Pim-1 negates the possibility of ligands interacting with residues other than the canonical hinge-ATP interaction, a hydrogen-bond acceptor interaction with the backbone carbonyl of E121. This dramatically reduces the potential affinity of hinge binders and is reflected in the higher K_m for ATP in PIM proteins, in the three-digit μM -range for Pim-1 vs single-digit

to double-digit μM for other Ser-Thr kinases [121]. Those developing Pim-1 ligands have consequently focussed on ATP-competitive rather than ATP-mimetic inhibition.

Far less common is the targeting of residues at the Specificity Surface (D128-D131). Staurosporine has been shown to bind in a ATP-mimetic fashion to the hinge and D128 via its methylamino group [122]. A High-Throughput Screen (HTS) and structure-guided optimisation campaign also resulted in very potent inhibitors that targeted the hinge, K67 and D128 via short but rigid chain linkers between ring groups [123,124] (Figure 3.4). But these are the only examples in the literature. This is perhaps surprising as these residues are conserved in all three PIM kinases and a focus of these campaigns was on the development of pan-PIM inhibitors.

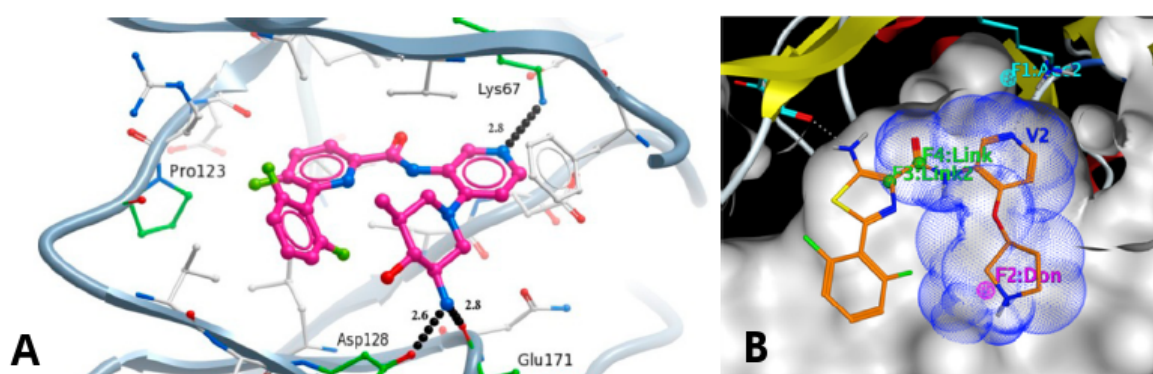


Figure 3.4: Example Pim-1 X-ray structures depicting rigid short-chain linkers between aromatic moieties interacting with polar residues in the binding pocket. (A) A ligand (pink carbons, ball-and-stick representation) is depicted forming polar interactions with residues (green, stick representation) such as Pim-1's catalytic Lysine (K67) and D128 (from Burger, et al 2013). (B) A ligand (orange carbons, stick representation) is depicted forming polar interactions with residues (light blue, stick representation) K67, the conserved backbone and a salt-bridge with D128 (from Ishchenko, et al 2015)

3.1.2 Structure-based Virtual Screening

In light of the structural information obtained from previous campaigns, Structure-Based Virtual Screening (SBVS) offers the possibility of large-scale structure-based ligand discovery for Pim-1. Large-scale docking campaigns have well-known limitations [9] and the size of drug-like chemical space is very large, estimated to be in excess of 10^{60} molecules [125]. Nonetheless, by fitting libraries of small molecules into a binding pocket and scoring them with a function containing well-developed physics-based terms and parameters, a relatively large amount of chemical space can be quickly sampled.

Fragment-based Drug Design (FBDD) offers the possibility of sampling a greater amount of this space than any HTS campaign [126]. This is considered a function of the greater

efficiency of exploring the chemical space of smaller molecules up to 30 heavy atoms ($\sim 10^{20}$ [127]). Fragments hits, despite lower potency overall, form high-quality, enthalpy-driven interactions with a target protein due a net loss of rigid-body entropy caused by the ligand binding event [44]. The small size of these molecules does, however, impact on detection of fragments. The use of X-ray crystallography has proven to be the most reliable method for finding ligands as many well-developed biophysical methods miss fragment-sized hits at their detection limits [46].

Target characteristics are also important in the design of a suitable library for FBDD. With respect to kinases, as they have evolved to bind ATP, molecules targeting the region where the adenine portion of ATP binds, known as ‘ATP-mimetic’, have proven successful at finding novel and selective ligands [128]. In contrast, some of the more unusual kinase hits have been developed to nM affinities from initial hits some distance away from the ATP binding site [129], or in an ‘ATP-competitive’ fashion, or ‘allosteric regulators’.

The synthetic feasibility of molecules from fragment hits has also been demonstrated to lead to ligands [130]. The pursuit of hits using methods to react chemical building blocks or fragments exhaustively *in silico* can, in principle, lead to easily-synthesisable molecules [57]. Ligands can be biased toward novelty, generated and applied to computational methods quickly [58]. Conversely, finding novel chemotypes is also possible with large and diverse libraries when compared to rational design campaigns from initial HTS hits. Rapid testing of molecules can also be facilitated by using libraries comprised of commercially-available molecules as ZINC Is Not Commercial (ZINC [131]).

The aims of this work are broadly within the scope of novel ligand discovery via SBVS and FBDD. In *Project 1*, fragments from an internal fragment library [45] were docked with Pim-1 and PKA and compared to X-ray crystallographic structures solved in-house. In an attempt to answer questions concerning which starting structure to use and how to develop reliable predictions, docking to multiple structures both simulated and derived from X-ray studies was conducted with multiple docking codes and results compared.

For *Project 2*, a library of extended fragments determined to produce reliable fragment binding modes was constructed via exhaustive *in silico* reactions, docked and tested against Pim-1 and a selectivity target, PKA. In *Project 3*, the same targets were screened against a lead-like subset of ZINC, tested in-house and by third-party vendor, protein-ligand complexes crystallised and their structures solved.

Finally, *Project 4* addresses an X-ray hit from the internal fragment library, Fragment number 200 (f200), that was not the complete molecule in the library but found a quality interaction with Pim-1's catalytic lysine and a nearby water network. The binding mode and protonation state of the fragment was confirmed by simulation, the fragment then extended *in silico* and docked, inhibitors then synthesised and tested.

3.2 Materials and methods

3.2.1 Protein preparation

Three-dimensional crystal structures of Pim-1 were prepared from in-house datasets or downloaded from the Protein Data Bank [59] (PDB: **3BGP**). Missing atoms or residues were repaired and minimised with Wit!P [132], hydrogens added and their positions minimised with CHARMM v31b2 [78]. Protonation states of histidine residues in all structures were determined via visual inspection. Each histidine was allocated a status of HID, HIE and HIP for docking dependent upon whether the ϵ , δ or both nitrogens in the imidazole side-chain were considered likely to be protonated.

Each protein structure was prepared for docking with three program suites: Solvation Energy for Exhaustive Docking (SEED) version 3.3.5 [60], OpenEye's OEDocking suite (v3.2.0.2) [69], DOCK 3.6 [68]/3.7 [24].

For SEED, defaults regarding departures from ideal hydrogen bonding geometry, clashes, and other chemical parameters were used and residues comprising the binding site of the protein were determined via manual inspection. Pre-defined rules regarding bond-length, angle and direction of H-bond donors and acceptors allow the construction and distribution of vectors of unitary length on all hydrogen bonding groups of ideal geometry. To reduce computations of unlikely polar-bond geometries, coordinates of a known ligand were used to restrict the number of force vectors evaluated to those not exceeding a set geometry criterion (10-70°).

The structure was prepared for docking with FRED using the *pdb2receptor* module in the OEDocking suite. Parameters regarding box shape, size and contours were calculated automatically and it was determined after preparation that the binding site was adequately covered.

Finally, the receptor was prepared for DOCK via the generation of a set of spheres representing the invagination of the receptor binding pocket. This was generated using a program

included in the DOCK distribution, *sphgen* [133]. The number and positions of spheres generated automatically was deemed adequate for coverage of the Pim-1 receptor. The oxygen atoms in E121 were ‘tarted’ by increasing their respective atomic partial charges by 0.4 with a consequent increase in the overall electrostatic potential surrounding this residue.

Images of the binding pocket for the SEED, FRED and DOCK binding pockets can be found in Appendix Figures 7.1, 7.2 and 7.3 respectively.

3.2.2 Ligand preparation

Molecules for all projects were drawn from an internal fragment library [45] as well as ZINC12 [131] and ZINC15 [14].

The internal fragment library was developed from databases of commercially available compounds with the focus on chemical novelty of ‘drug-like’ fragments amenable to development strategies such as growing and merging. This process resulted in a fragment library of 361 molecules that was validated against an aspartic protease receptor, Endothiapepsin [48].

For docking with SEED and FRED, the fragment library was prepared from Simplified Molecular Input Line Entry System (SMILES [134–136]) using OpenEye’s QUACPAC [137] suite for appropriate protomer and tautomer generation. Multiple 3D conformers for each protomer were generated with OMEGA [138]. To reduce the number of similar conformers and lower the computational effort required, values of the *-rmsrange* parameter were increased sequentially from 0.5 - 1.2 with number of rotatable bonds (1-7) in a given molecule. The *-erange* parameter was set to 5.0 kcal·mol⁻¹. This reduced the number of conformers from hundreds to < 20 in most cases without substantially compromising the amount of conformational space explored. AM1-BCC [139] partial atomic charges were calculated for each conformer.

Generation of the fragment library for use with DOCK was completed using the internal ligand generation pipeline implemented with the DOCK 3.7 suite to generate the .db2 files required. The pipeline takes SMILES as input and uses ChemAxon [140] programs for protomer generation, CORINA 4.1 [141] for initial 3D conformer generation, AMSOL 7.1 [142] to compute atomic partial charges and ligand desolvation terms and finally OMEGA to generate conformers.

Both ZINC12 and ZINC15 are freely-available databases of, in principle, biologically-relevant and purchasable compounds. Additional web-based resources are available such as similarity

searches and activity prediction. The database focusses on chemical novelty and offers several ‘tranches’ for download such as ‘fragment’, ‘lead-like’, ‘drug-like’, etc. For this work, only lead-like databases were downloaded directly from the ZINC12 and ZINC15 websites.

Finally, after initial fragment docking, extensions to the internal fragment library were generated using the Python *IN silico* de novo Growing Utilities (PINGUI [58]). This method takes a given molecule, identifies sites, or, ‘reactive handles’ on the molecule amenable to extension and reacts them exhaustively *in silico* with building-block libraries using 58 robust reactions [57] to ensure synthetic feasibility.

Molecules in another series of aniline triazoles were reacted *in silico* with a fragment derived from an in-house developed database, SCUBIDOO [55]. The scaffold (**4012414**) and all synthesised derivatives were predicted to interact with both K67 and the hinge (E121). A coupling reaction, Buchwald-Hartwig cross-coupling, was chosen to extend the molecules at the *meta* position of the aniline in order to target D128. In total, 6 molecules were selected from this series for docking and testing (Figure 3.5).

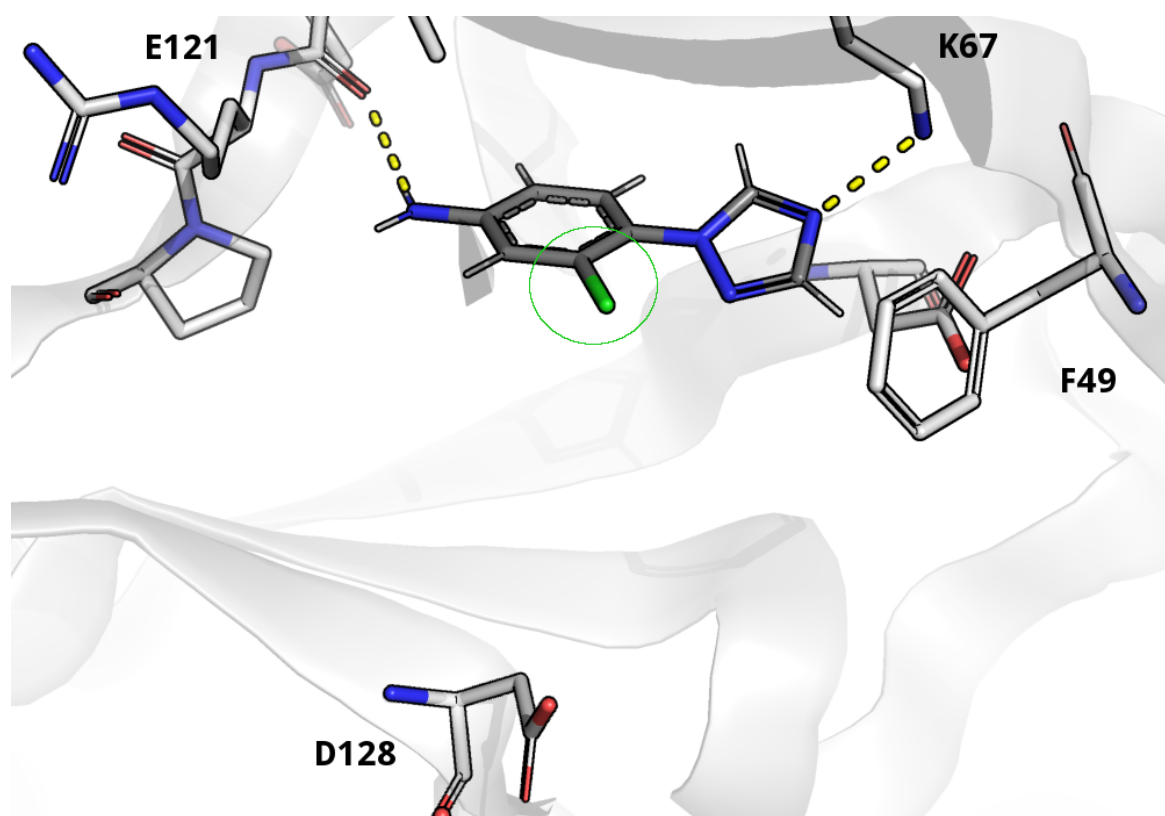


Figure 3.5: Depicted here is the Pim-1 binding pocket in complex with an aniline triazole scaffold (dark grey, stick representation) and key residues (light grey, stick representation) labelled. Predicted polar interactions are represented by yellow dashed lines. The scaffold’s triazole was predicted to interact with the catalytic lysine (K67) and its aniline nitrogen predicted to interact with a hinge residue (E121). A halogen in *meta* position of the aniline is highlighted (green circle) to show the proposed *in silico* reaction site

For *Project 4*, extensions to a ligand found for Pim-1 from the internal fragment library (*f200*) were generated at position 5 on the aryl component of its quinoxaline core (Figure 3.6). Extensions from this position therefore made amenable the targeting of charged residues near the Specificity Surface (D128 and D131). One coupling reaction was used (Suzuki [143]) and this was reacted *in silico* with building blocks from the MolPort [144] library.

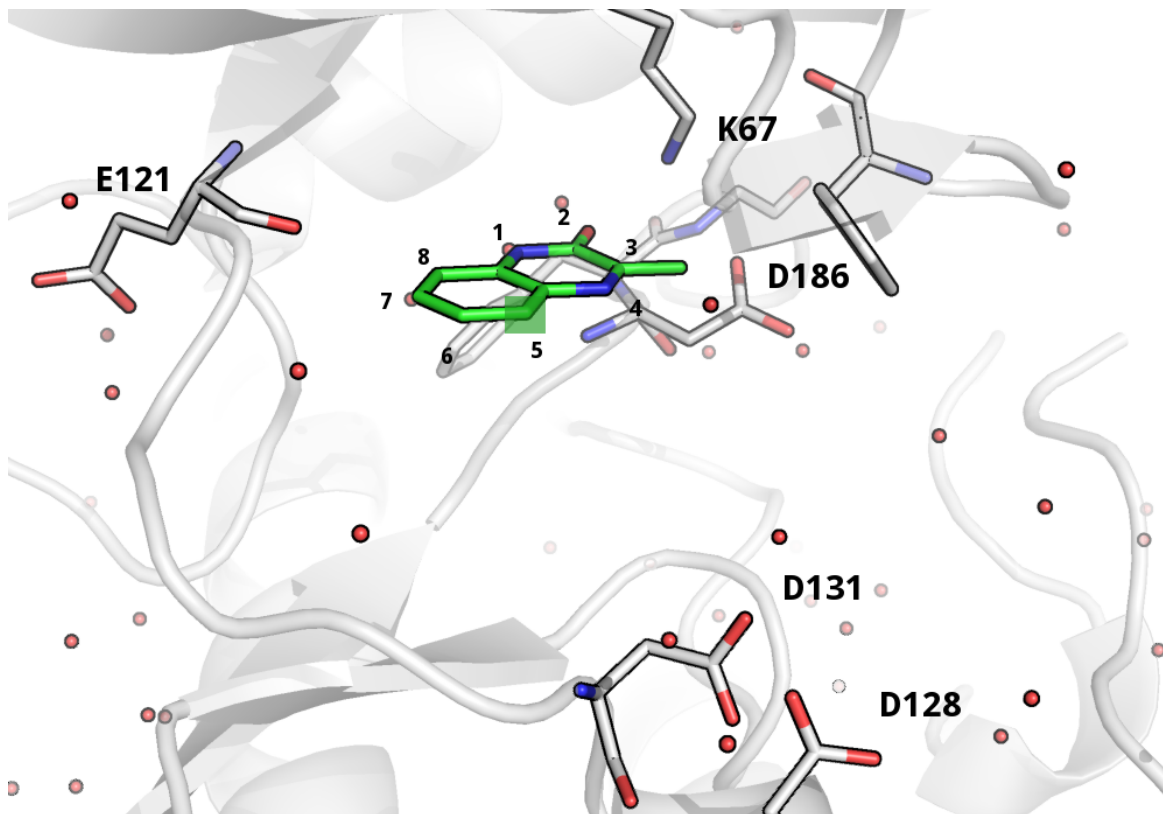


Figure 3.6: Depicted here is the Pim-1 binding pocket in complex with a substituted quinoxaline (green carbons, stick representation) and key residues (light grey, stick representation) labelled. Position 5 of the quinoxaline is highlighted (green square) to show the proposed *in silico* Suzuki reaction site

3.2.3 Molecular dynamics simulations

All molecular dynamics simulations were completed with the Amber and AmberTools suite of programs (versions 14 [145] and 16 [146]). Simulations post-processed with MDMix [19] were completed with Amber 16.

Ligands were prepared from files containing atomic coordinates (*.mol2*) and fixed partial charges calculated using the Restrained ElectroStatic Potential (RESP) model [147]. This was completed via geometric optimisation and electrostatic potential calculations on converged wave functions using Gaussian 09 rev C.01 [148] at the HF/6-31G* level of theory. Software in AmberTools (*Antechamber* [149]) was used for file conversion between Gaussian logs and *.mol2* files.

Protein structures were initially prepared using MOE [150] to add hydrogens, fix chain breaks and missing atoms. The positions of all crystallographic water molecules were retained. Amber Tool's LEaP module was used to generate topologies and coordinates for both ligands and proteins. The force fields chosen for all simulation steps (heating, equilibration and production) were The General Amber Force Field (GAFF [76]) for ligands and ff99SB [151] for the protein, respectively. The protein-ligand complex then had charges neutralised with Na^+ and Cl^- ions and was solvated with a truncated-octahedral TIP3 [152] water box of 10 Å in size repeated in all three spatial dimensions such that the closest distance between any atom present in the protein and the edge of the periodic box was also 10 Å to ensure total coverage of the complex.

Prior to production MD, each complex first underwent two energy minimisation stages, then a heating stage from 100K to 300K and six subsequent equilibration stages at constant volume and temperature for a total of 5 ns (2 fs time steps). All heating and equilibration stages had protein heavy atoms restrained with restraint weights progressively reduced to zero as the complex progressed from one stage to the next.

Upon confirmation of a complex's thermodynamic stability in terms of temperature, pressure potential energy and other variables, a constant volume and temperature (NVT) simulation was conducted with periodic boundary conditions. All bonds with hydrogen were constrained using the SHAKE algorithm [153] and Langevin thermostat [154] for temperature regulation. After checks for stability of the system (e.g. temperature, pressure, RMSD), each production simulation was run for 100ns with three repeats per simulation for a total of 400ns of simulation time.

For Project 1, the Hierarchical Agglomerative Clustering algorithm in AmberTools' CPP-TRAJ was used to output the most frequent conformation of the protein binding pocket. Five clusters were calculated and the average distance between members of clusters was used to determine membership of each cluster.

For the purposes of trajectory analysis, a hydrogen bond between two atoms is defined as between 2.4-3.4 Å heavy atom distance and angle of 120-240°.

MDMix simulations in Project 4 were conducted on the Pim-1 apo structure. These simulations used an additional protein forcefield, ff99SB-ILDN [155], that cites improved side-chain potentials for the ff99SB force field.

Each simulation had a single minimisation stage and five equilibration stages where the temperature was progressively increased to 300K at constant pressure and restraints were only applied to non-hydrogen atoms. Simulations also used a 20% mixture of three solvents in a truncated octahedral box; water, acetamide and ethanol (Figure 3.7). Each production simulation was run for 50 ns at constant pressure (NPT) with four replicates making a total of 15 simulations (3 solvents x 5 simulations per solvent), or, 250 ns of simulation time per solvent.

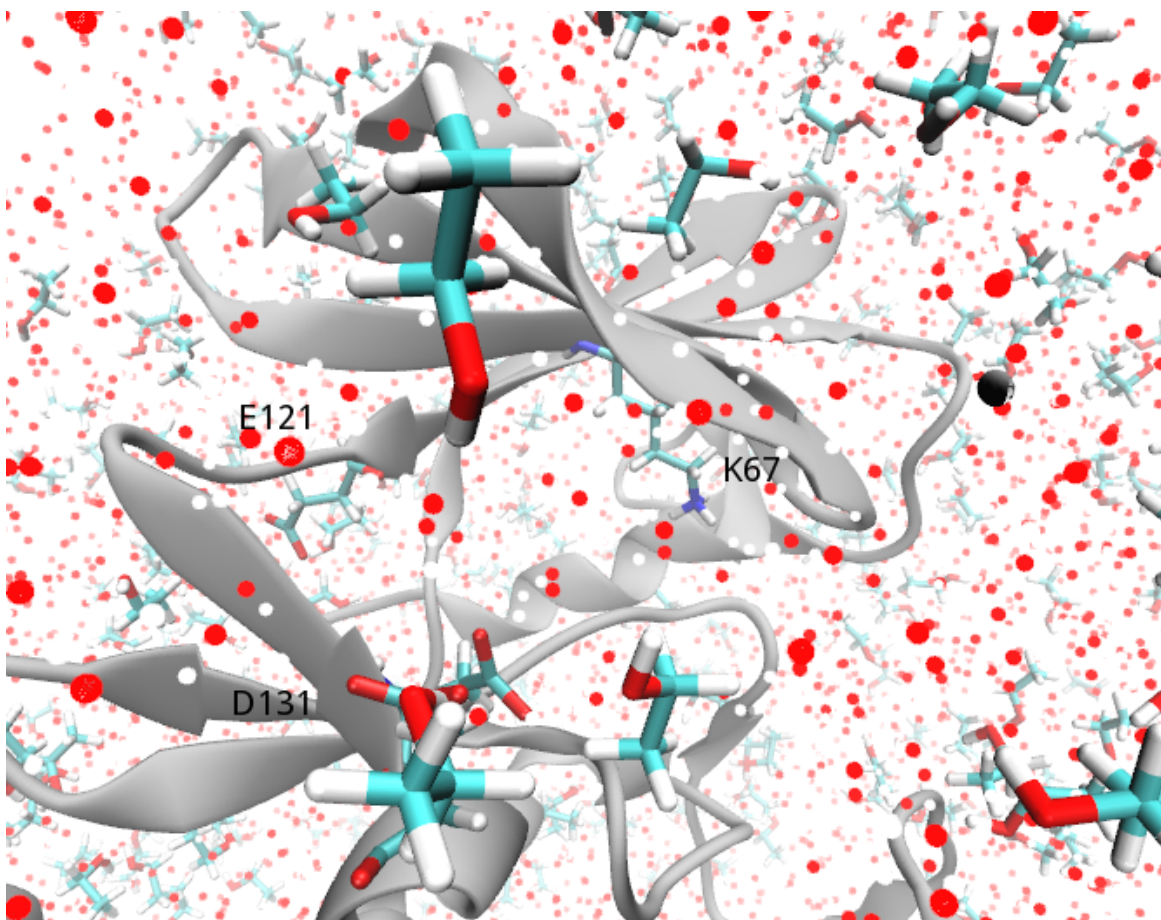


Figure 3.7: Example Pim-1 apo structure with 20% ethanol/water solvent mixture and binding pocket residues labelled. Red spheres are water molecules whereas ethanol molecules are depicted as stick representations

From these simulations, binding ‘hot spots’ were calculated from solvent counts per 5 \AA^3 grid element. The observed vs expected population was used to calculate the binding free energy using the inverse Boltzmann relation.

$$\Delta G_{\text{binding}} = -k_B T \cdot \ln\left(\frac{N_i}{N_0}\right) \quad (3.1)$$

Solvent densities were visualised as isomeses of $\Delta G \leq -1.0 \text{ kcal}\cdot\text{mol}^{-1}$.

3.2.4 Docking and minimisation

For FRED and HYBRID, program defaults were used with the exception of *dock_resolution* (=HIGH). Only the highest scoring pose for each molecule was retrieved.

For SEED, both polar and apolar docking algorithms were selected and all poses to an evaluated energy of $+5.0 \text{ kcal}\cdot\text{mol}^{-1}$ were retrieved. Otherwise, program defaults regarding protein clashes, charge cut-offs and so on were used. Poses were assessed in terms of both fragment and receptor desolvation energies and total interaction energies.

For DOCK, parameters governing van der Waals clashes were increased from defaults ($20 \text{ kcal}\cdot\text{mol}^{-1} \rightarrow 500 \text{ kcal}\cdot\text{mol}^{-1}$) i.e. poses where the vdW score exceeded this value were discarded. The *matchgoal* parameter was also set higher ($1000 \rightarrow 5000$) to increase sampling of molecular orientations. Simplex minimisation was used. Sphere colouring was not used. Only the highest scored pose was retrieved. Regarding all other parameters, program defaults were used.

Protein structures from simulations often results in residues where side chain rings and backbone amides were bent significantly out-of-plane. To alleviate this, atomic positions were energy minimised using CHARMM with different harmonic force constants applied at different levels of the protein. Force constants were applied by atomic mass resulting in a force on a given atom in $\text{kcal}\cdot\text{mol}^{-1}/\text{\AA}^{-1} \cdot 2$.

Prior to docking, simulated structures were energy-minimised with CHARMM, progressively weaker harmonic restraints applied at different levels of the protein. Protein backbone atoms were the most restrained with a force constant of 2.0, side chain atoms at 0.4 with hydrogens the most unrestrained at 0.08.

The positions of all pose structures within the Pim-1 active site were optimised using SZYBKI [156]. Optimisation parameters were set to minimise torsional strain of molecular poses using the Merck Molecular Force Field (MMFF94 [157]) with the Poisson-Boltzmann solvation model [158]. All residues within 3 \AA of a given pose were also minimised. Program defaults regarding the optimisation method (e.g. BFGS [159] algorithm) were used.

To measure the quality of a docking prediction with respect to crystallographically-determined atomic positions, the Root Mean Square Deviation (RMSD) of atomic positions was used. It is simply the quadratic mean of the distance δ between equivalent atoms N in two superposed chemical structures.

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N \delta_i^2} \quad (3.2)$$

Molecular poses retrieved less than 1 Å RMSD from the ligand’s crystal position were considered a near-perfect fit with poses less than 2 Å considered very good and less than 3 Å to be good.

Clustering of poses for *Project 2* was completed with ChemMine [160], an online service for small molecule data analysis. A Tanimoto similarity (T_c) of cut-off of 0.4 or lower using the Extended-Connectivity Finger Print version 4 fingerprint (ECFP4 [94]) was used with their hierarchical clustering functionality to sort molecules into ‘bins’. The first compound from each bin was retained.

All similarity checks of molecules were conducted against the ChEMBL [12] database at <https://www.ebi.ac.uk/chembl/>.

Assays

In Project 2 potential inhibitors were tested against Pim-1 and a selectivity target, Protein Kinase A (PKA) using a *Z'-LYTETM Kinase Assay Kit* (Invitrogen/Thermo-Fisher) with substrate peptides appropriate for the protein used.

The peptide substrate is labelled with two fluorophores (coumarin and fluorescein) that make up a Fluorescence Resonance Energy Transfer (FRET) pair and in the primary reaction, the γ -phosphate of ATP is transferred to a serine or threonine residue on the peptide. In subsequent reactions, a site-specific protease cleaves unphosphorylated peptides, disrupting FRET between donor (coumarin) and acceptor (fluorescein) fluorophores (Figure 3.8). From here, an emission ratio of donor vs acceptor emissions after excitation at 400 nm is calculated.

Both cleaved and uncleaved FRET-peptides contribute to the fluorescence signals. The emission ratio will therefore remain low if the peptide is phosphorylated (i.e. low kinase inhibition) and will be high for unphosphorylated peptide, reflecting inhibitory activity.

For both enzymes used in the assay, an initial optimisation step was performed to determine the appropriate enzyme concentration for screening. The assay protocol recommends near- K_m , ATP concentrations and kinase concentration that phosphorylates 20-40% of the assay substrate peptide.

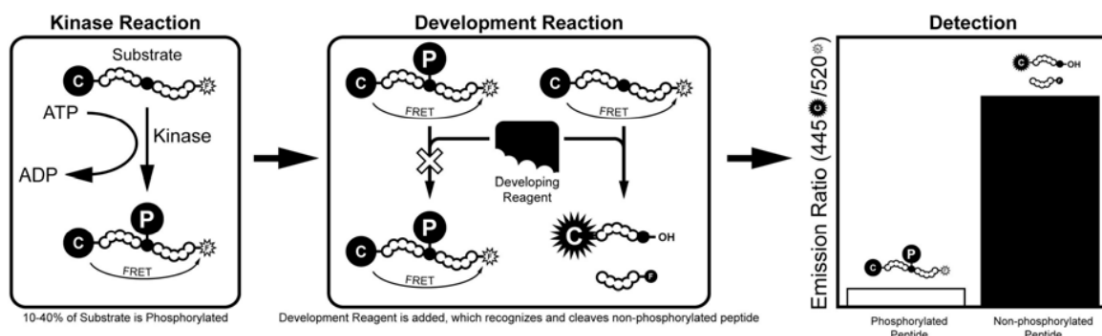


Figure 3.8: Z'LYTE biochemical assay schematic depicting the (1) Kinase Reaction, where a synthetic peptide is phosphorylated by the kinase; (2) the Development Reaction, where unphosphorylated peptides are cleaved by the Development Reagent and; (3) measurement of emission ratios of cleaved vs uncleaved peptide reflecting the presence (or absence) of fluorophores in FRET and, therefore, inhibitory activity (from Z'-LYTE KINASE ASSAY KIT – SER/THR 7 PEPTIDE PROTOCOL, <https://www.thermofisher.com/order/catalog/product/PV3180>)

For both Pim-1 and PKA, the kinase concentration meeting these requirements was determined to be 80 µg/mL (20 µg/mL final concentration in each well based on a 10 µL assay). The ATP concentration used was 16 µM (4 µM final concentration) with K_m determined to be in the range of ~4.2 µM with each assay. Inhibitors were initially solved in 4% DMSO and serially diluted to one-digit nM concentration and a final DMSO concentration in each well of 1%. Staurosporine was tested for comparison purposes in each assay performed.

All compounds were tested with two mechanical and three biological replicates in black, flat-bottom 384-well plates (Greiner MODEL; 7849000) and FRET of both *coumarin* (445nm) and *fluorescein* (520nm) fluorophores was measured at each well and the emission ratio calculated.

Per cent phosphorylation was calculated as a function of the emission ratio of compounds tested and controls thusly:

$$\% \text{ Phosphorylation} = 1 - \frac{(EmissionRatio \times F_{100\%}) - C_{100\%}}{(C_{0\%} - C_{100\%} + [EmissionRatio \times (F_{100\%} - F_{0\%})])} \quad (3.3)$$

All IC_{50} plots were created in Qtiplot [161]. Dose-dependent inhibition was then quantified by plotting fitting a sigmoidal curve via a 4-parameter logistic equation [162]. The IC_{50} of a given compound was calculated at the mid-point of the curve, indicating half-maximal inhibition of the protein by a given inhibitor.

The binding affinity K_i of an inhibitor in a binding assay of fixed substrate concentration $[S]$ and half-maximal substrate enzyme activity K_m was from the obtained IC_{50} 's using the Cheng-Prusoff relation [163].

$$K_i = \frac{IC_{50}}{1 + \frac{[S]}{K_m}} \quad (3.4)$$

The Hill slope (d) of the plot was used as an indicator of possible ligand aggregation [164].

$$y = min + \frac{max - min}{1 + (\frac{x}{IC_{50}})^d} \quad (3.5)$$

For Project 3, compounds were tested both in-house and by a third-party vendor (Eurofins Inc.). Compounds were tested against Pim-1 and their residual activities at a concentration of 10 μ M calculated via single-point inhibition assay [97] against a panel of human kinases. Compounds displaying activity were then re-tested at concentrations close to their likely IC_{50} .

Inhibition constants K_i were determined of molecules with measured IC_{50} s using the Cheng-Prusoff relation [165]. We assume Michaelis–Menten kinetics of all enzymes tested, interacting with a substrate of Michaelis constant K_M and concentration S .

$$K_i = \frac{IC_{50}}{1 + \frac{[s]}{K_m}} \quad (3.6)$$

3.3 Results

3.3.1 Project 1: Docking software comparisons vs structures from MD simulation

Three X-ray structures were chosen for comparison, one from the PDB (3BGP) and two resulting from equilibrium simulations of the apo structure (apo_MD) and 3BGP (3BGP_MD).

The 3BGP structure was chosen as it most broadly represented structural characteristics thought important for ligands binding to Pim-1. It did not differ markedly from 97 other Pim-1 structures on the PDB (RMSD $\bar{x} = 0.37 \text{ \AA}$, $\sigma_{\bar{x}} = 0.05$) and was therefore chosen for further docking.

Trajectories from simulations of the apo_MD structure were combined for cluster analysis. The first frame in the trajectory was chosen as the reference frame and, after alignment, RMSDs calculated for binding pocket residues, K67, the P-Loop and the DFG loop. It was determined that the overall RMSD of the protein did not vary markedly during the course of the simulation and quickly achieved an equilibrium ($<1 \text{ ns}$).

The first cluster for each of the 3BGP_MD and apo_MD binding pockets contained 64% and ~86% of the total frames from trajectories respectively. There was overall a greater variation in RMSDs between each frame and the centroid of each cluster for the 3BGP_MD binding pocket (RMSD \bar{x} = 2.10 Å) when compared to the apo_MD pocket (RMSD \bar{x} = 1.83 Å). This variation was also reflected in average RMSDs from cluster centroids when the P-Loop alone was considered (RMSD \bar{x} = 2.44 Å for 3BGP_MD vs 1.38 Å for apo_MD). This variation in-turn reflects greater flexibility of the P-Loop in the presence of a bound ligand as, presumably, the absence of a ligand allows flexible residues to return to a more relaxed state.

However, with respect to the 3BGP_MD trajectory, the first cluster was closest in terms of RMSD (0.861 Å) to its crystal structure when compared to the other four clusters (RMSD \bar{x} = 1.30 Å), reflecting a more 'averaged' bound ligand state and was thus chosen for fragment docking.

Descriptive statistics regarding simulations of this structure can be found in the Appendix Table 7.1.

A Thermal Shift Assay (TSA) was performed as a crystallographic pre-screening step on all 361 compounds in the internal fragment library and from these, 31 molecules displayed a 'significant' (> 4 °C) positive temperature shift, indicative of stability in the protein-ligand complex (Appendix, Figure 7.4). The crystal structures of 12 of these molecules in complex with Pim-1 were subsequently fully resolved and used as references for assessing docking performance.

All 12 fragments were docked to the the 3BGP X-ray structure as well as the apo_MD and 3BGP_MD structures. When docking to the X-ray structure of 3BGP, SEED recreated polar contacts in most (92%) cases and retrieved at least one pose per fragment close to the crystal binding mode (Table 3.1). FRED and HYBRID performed similarly when compared to each other, in over half of cases retrieving poses close to the crystal binding mode of each fragment. DOCK was unable to retrieve poses near the crystal binding mode.

Broadly improved pose retrieval was achieved when docking to the apo_MD structure. SEED and HYBRID were able to retrieve more poses that recreated the crystal binding mode but SEED, FRED and HYBRID all performed worse in terms of RMSD. Improved performance in terms of the number of crystal poses retrieved and RMSD was noted with DOCK.

Docking fragments to the 3BGP_MD structure improved the RMSD of fragments from the crystal binding mode for both HYBRID and SEED with respect to apo_MD structure whilst also achieving similar performance to the 3BGP crystal structure. Unlike, the 3BGP crystal structure, poses were retrieved for DOCK when using the 3BGP_MD structure for docking.

Program	PDB:3BGP		apo_MD		3BGP_MD	
	RMSD \bar{x} ($\sigma_{\bar{x}}$)	% poses	RMSD \bar{x} ($\sigma_{\bar{x}}$)	% poses	RMSD \bar{x} ($\sigma_{\bar{x}}$)	% poses
FRED	2.58 (1.33)	75%	3.07 (1.15)	50%	3.45 (1.06)	58%
HYBRID	2.61 (1.17)	67%	3.37 (0.97)	75%	2.69 (1.38)	58%
SEED	1.68 (0.50)	92%	2.62 (0.95)	100%	2.15 (1.15)	100%
DOCK	- (-)	-	3.98 (1.09)	50%	3.90 (0.64)	42%

Table 3.1: Comparison of mean and standard deviations of RMSDs (\AA) from fragment X-ray structures for poses docked to PDB:3BGP, apo_MD and 3BGP_MD. Per-cent poses is defined as the proportion of docked poses within 5 \AA of the position of the bound X-ray ligand

Complete data regarding fragment RMSDs from X-ray crystal structures against both 3BGP and apo_MD targets can be found in Appendix Tables 7.2 and 7.3 respectively.

3.3.2 Project 2: synthetically-feasible extensions of an internal fragment library

The initial fragment project found 54 fragments with a stable binding mode in Pim-1 of which 34 possessed so-called ‘reactive handles’ amenable to extension. A thermal-shift assay was conducted and 20 (59%) displayed a positive temperature shift, indicative of a net increase in stability of these protein-ligand complexes. From these fragments, 37 reactions *in silico* with 7805 building blocks from the ChemBridge [56] dataset generated a docking library of 171 919 (313 299 protomers).

As there were large disparities in the number of molecules generated by each *in silico* reaction, the full library was divided into separate docking instances by reaction to avoid ‘drowning out’ of potentially good but similarly scored molecules from each reaction. Where there were 500 or more poses retrieved, the top-500 from each dock was retained for analysis. Otherwise, each pose in a given dock was analysed.

A complete list of each dock and the size of each library can be found in the Appendix Table 7.5.

Visual inspection of poses retrieved resulted in library of 1156 molecules that was reduced to 517 after minimisation and eliminating those with unsatisfactory binding modes (e.g.

stranded H-bond donors, shape complementarity). They were divided into groups depending on whether they retained the binding mode of the fragment from which they were derived (394) or not (123). Each of these groups was further divided into separate groups depending on whether they interacted with residues near the Specificity Surface or not.

In the case of molecules that retained their fragment binding mode, 146 poses displaying an interaction with D128 were retrieved whilst 248 poses were not predicted to be interacting with this residue. Of molecules displaying a different pose to their original fragment binding mode, 20 molecules interacting with D128 were retrieved whereas 103 were not predicted to be interacting with this residue.

There were many more molecules retaining their predicted fragment binding modes as compared to those where the extended molecules displayed a different binding mode from the original fragment. ECFP4 Tanimoto similarity was used to cluster this group into ‘bins’ and the first molecule from each bin retained. After clustering and visual inspection, common in high-throughput and virtual screening [166,167], a final list of 47 molecules was collated. This list comprised of 27 molecules displaying a binding mode extended from their originally predicted fragment position (‘BM’) and 20 where the extended molecule displayed a different binding mode to the initial fragment prediction (‘novel’). These molecules were subjected to assessment and grading by lab members experienced in docking, the grades collected and analysed with 21 chosen for synthesis. Derivatisation and optimisation of some compounds resulted in a final list of 37 compounds (Figure 3.9).

Whilst there was a great diversity in which moieties were positioned within the Pim-1 pocket, the binding mode for all predictions was similar. All poses retrieved were interacting with K67 and most were also predicted to be interacting with residues near the Specificity Surface, particularly D128. In most cases, interactions with K67 were via a salt-bridge with a carboxylate whereas D128 was generally predicted to interact with a charged endocyclic amine group such as a protonated piperidine (Figure 3.10). Due to the bi-partite nature of molecules created by PINGUI, most molecules were comprised of two or more ring systems with flexible linkers rather than the bulky π -systems of known ligands for Pim-1 that interact with D128 e.g. Staurosporine, Bisindolyl Maleimides.

A complete list of molecules **13** to **49** can be found in the Appendix Table 7.6.

It was decided from the above predictions to synthesise 10 compounds. Of these, 9 were actually synthesised with one failing (Table 3.2). All were racemic mixtures. Of these, 3

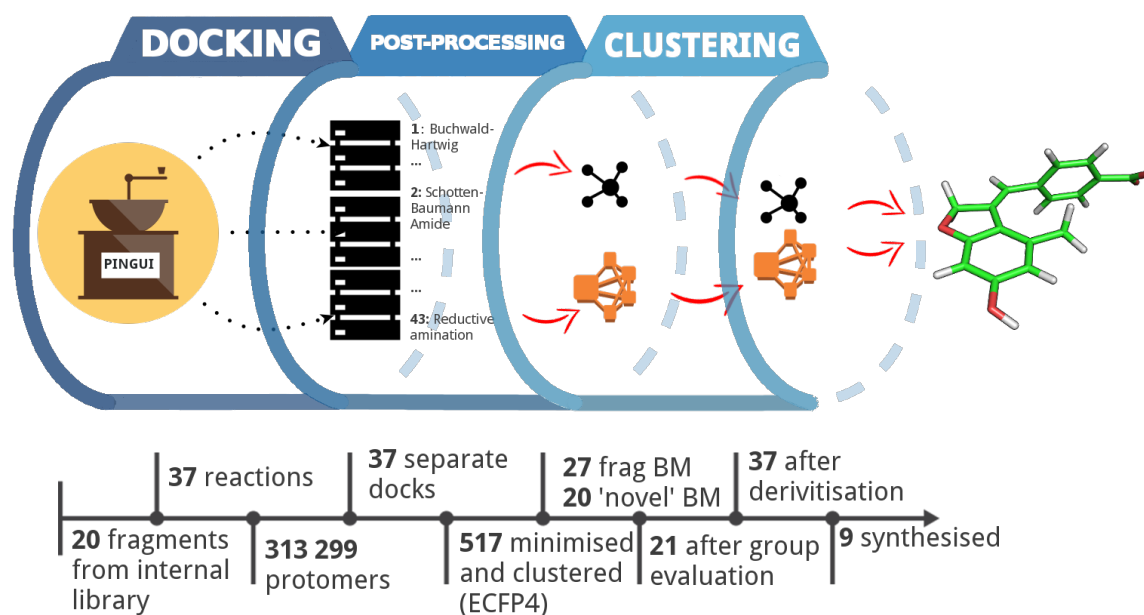


Figure 3.9: Schematic of Project 2 preparation, docking and post-processing. Docked molecules were first created by reacting internal library fragments *in silico* with ChemBridge building blocks. They were then divided into separate docking instances by reaction type, docked and the results aggregated and clustered by Tanimoto similarity (ECFP4 fingerprint). After detailed visual inspection, groups were formed based upon whether molecules retained the original fragment binding mode or not (frag BM and 'novel' BM respectively). A preliminary list in each of these categories was subjected to visual inspection and grading by lab members prior to the formation of a final list for analysis by collaborators for synthesis. Minor alterations and synthetically-feasible derivatives were suggested by collaborators prior to the final synthesis of 9 molecules that were then tested for inhibitory activity via FRET-based assays.

molecules (**27**, **29** and **47**) were the original molecules predicted to interact with Pim-1 and **6** were synthesised with small alterations (e.g. pyridine → benzene) to ensure synthetic feasibility. After minimisation, all 9 molecules were predicted to have net favourable interaction energies (Appendix Tables 7.7, 7.8 and 7.9)

All molecules were tested via FRET assay with **5** showing at least mild activity against Pim-1. The most potent was **26**, measured to be inhibiting at an $K_i=24.99\mu\text{M}$. This is as compared to Staurosporine, measured to be inhibiting at a calculated $K_i=4.7\mu\text{M}$ (Figure 3.11).

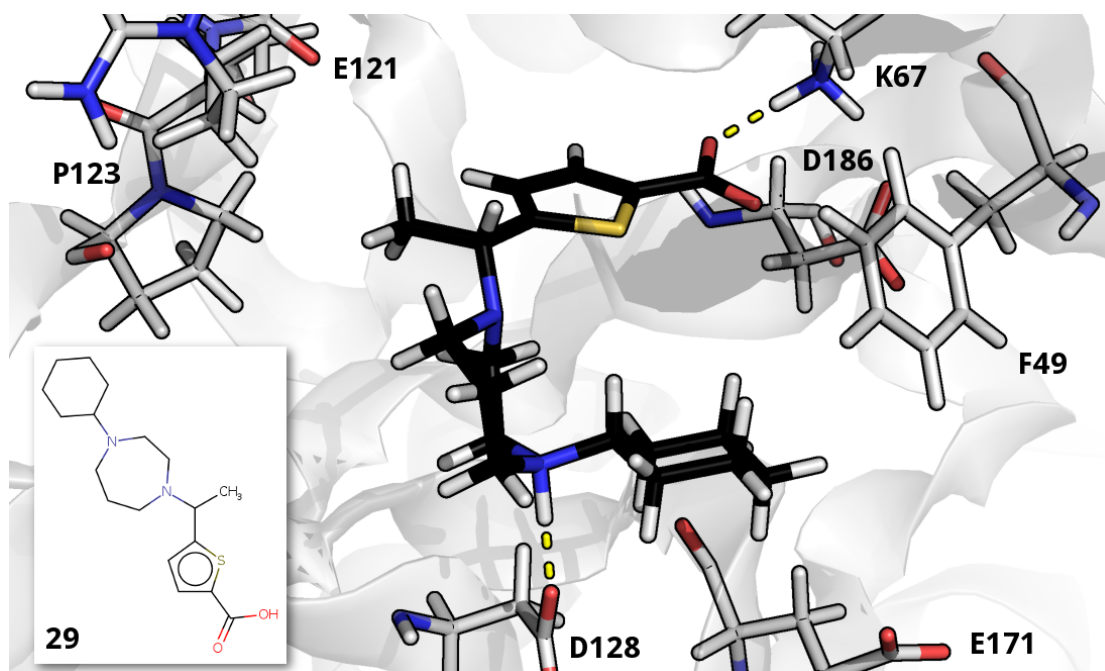


Figure 3.10: Predicted binding mode of **29** (black carbons, stick representation) in the Pim-1 binding pocket, key residues (light grey, stick representation) labelled and 2D structure (inset) of the pose displayed. Predicted polar interactions are represented by yellow dashed lines. Depicted here are salt-bridge interactions formed between **29**'s carboxylate and K67 as well its diazepane nitrogen and D128

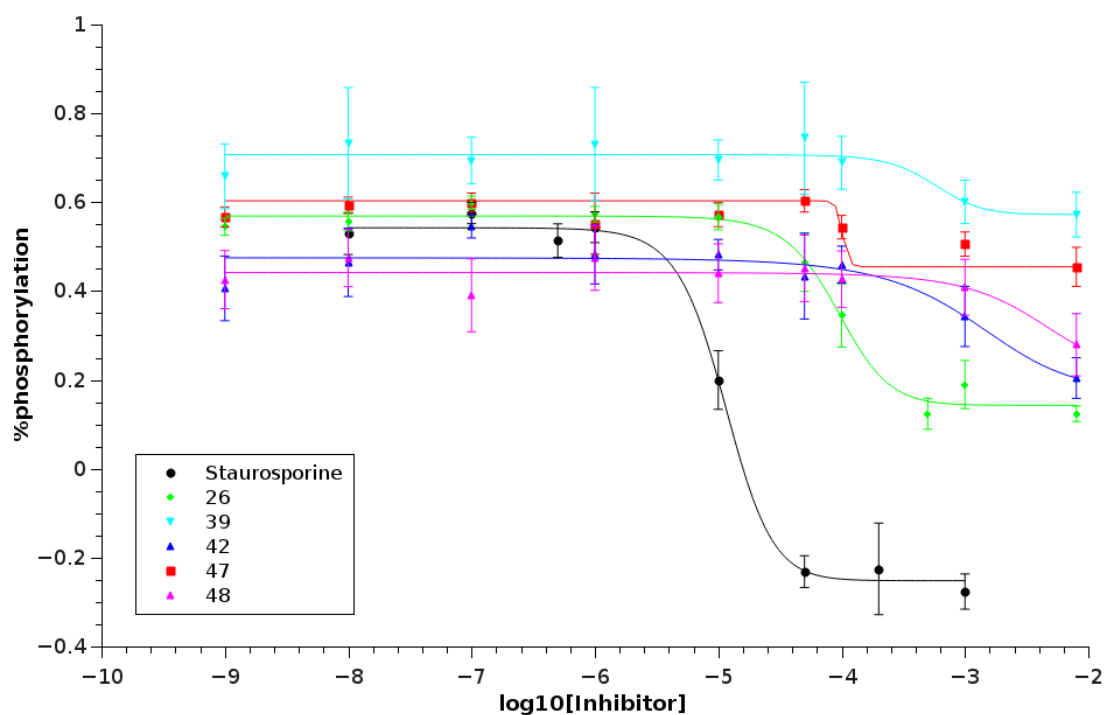


Figure 3.11: Dose-response curves of FRET-based assays from extensions via *in silico* reactions with our internal library. Depicted here is the per-cent phosphorylation against the LOG_{10} ligand concentration of tested compounds displaying activity against Pim-1 with error bars (standard-error of measurement) and curves fitted to the data (multi-coloured lines). The IC_{50} of a given compound was calculated at the mid-point of the curve, indicating half-maximal inhibition of the protein. The most potent ligand, **26**, is represented by the lime green line with Staurosporine (black line) for comparison. Each experiment was repeated with two mechanical replicates and three biological replicates.

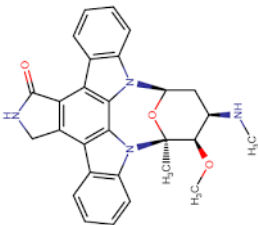
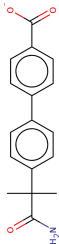
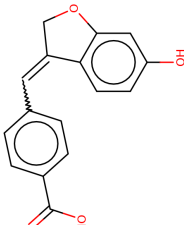
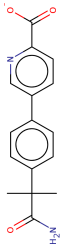
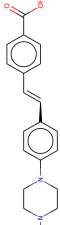
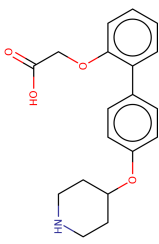
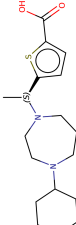
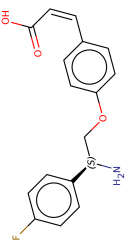
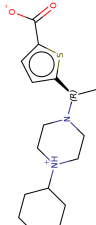
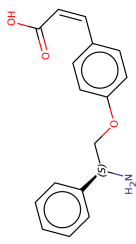
ID	K _i (μM)	2D Structure	ID	K _i (μM)	2D Structure
Staurosporine	4.71 μM		48	250.44 μM	
26	24.98 μM		47	191.76 μM	
27	>10mM		42	91.47 μM	
29	>10mM		40	>10mM	
30	>10mM		39	141.14 μM	

Table 3.2: All 9 compounds tested from extensions via *in silico* reactions with the internal library, their measured K_i and 2D structure. The measured K_i of Staurosporine is included for reference. Each experiment was repeated with two mechanical replicates and three biological replicates.

Compound **26** was predicted to interact with the side chains of K67 and D128 and the backbone of D186 (Figure 3.12).

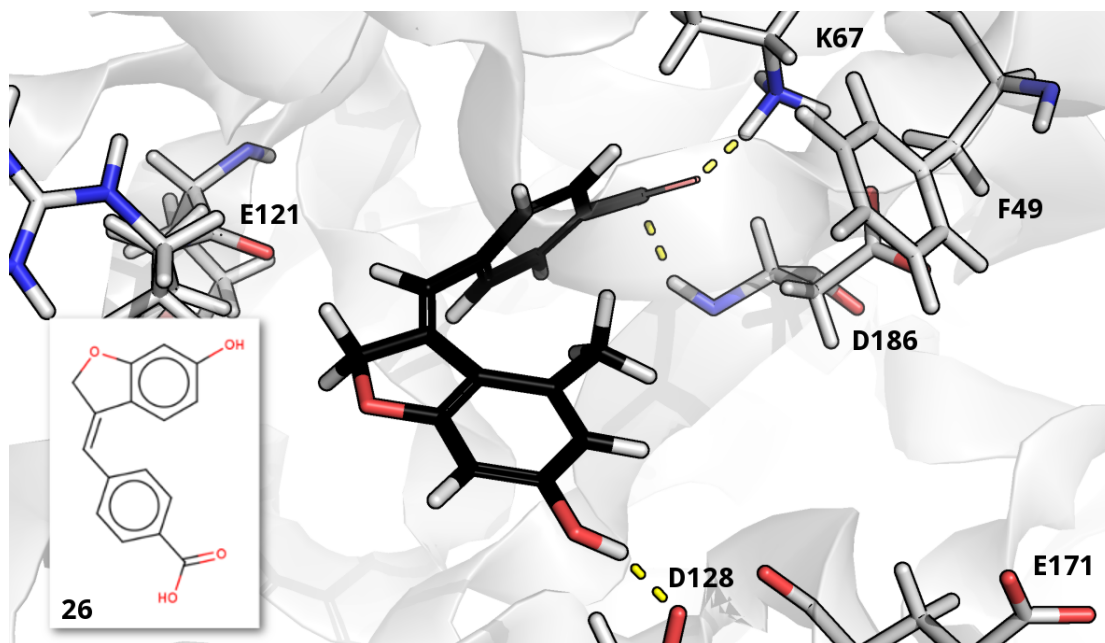


Figure 3.12: Predicted binding mode of the (E)-isomer of **26** (black carbons, stick representation) in the Pim-1 binding pocket, key residues (light grey, stick representation) labelled and 2D structure (inset) of the pose displayed. Predicted polar interactions are represented by yellow dashed lines. Depicted here is a salt-bridge interaction between the carboxylate of **26** and K67 with the hydroxy group of its substituted benzofuran interacting with D128. This ligand was the most potent of all tested with a calculated K_i of 24.98 μM .

All molecules from the aniline triazole series were tested via FRET assay with **52** the only ligand to result in a measurable K_i of 137.0 μM (Figure 3.13). However, although a curve could be fit, the values of per-cent inhibition suggest very little inhibition by this compound was actually occurring.

Molecule **52** was the only ligand in the series predicted to directly target D128 (Figure 3.14).

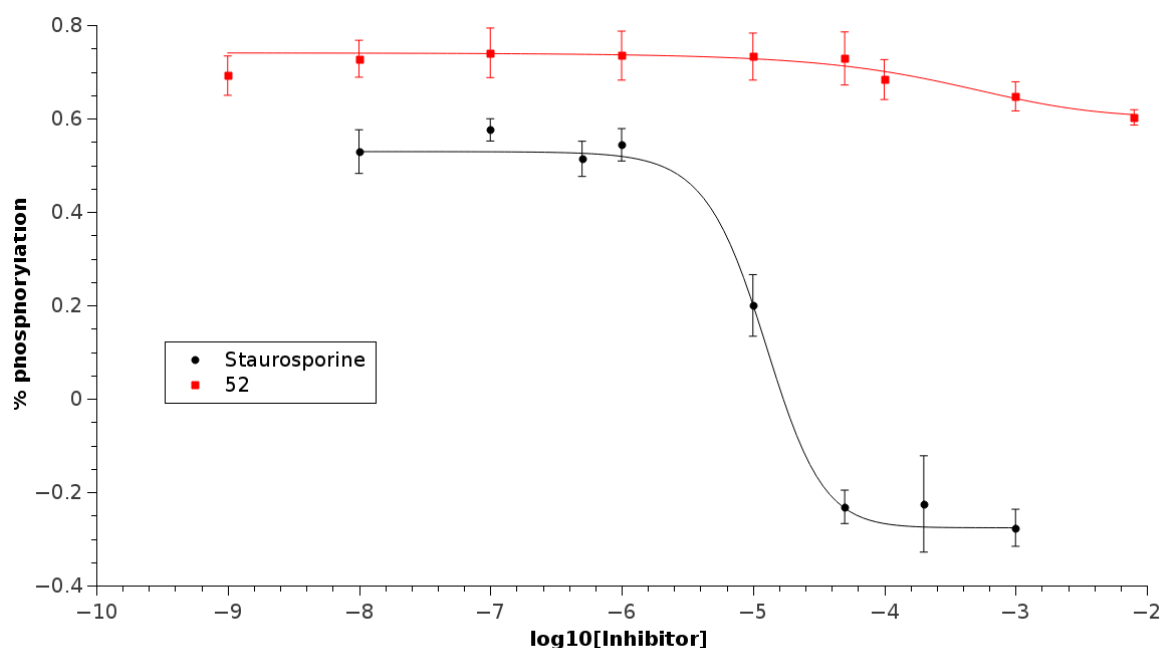


Figure 3.13: Dose-response curves of FRET-based assays from extensions via *in silico* reactions with our internal library. Depicted here is the per-cent phosphorylation against the LOG_{10} ligand concentration of tested compounds displaying activity against Pim-1 with error bars (standard-error of measurement) and curves fitted to the data (multi-coloured lines). The IC_{50} of a given compound was calculated at the mid-point of the curve, indicating half-maximal inhibition of the protein. In this case, **52** (red line) is displaying very mild inhibition. Staurosporine (black line) is again depicted for comparison. Each experiment was repeated with two mechanical replicates and three biological replicates.

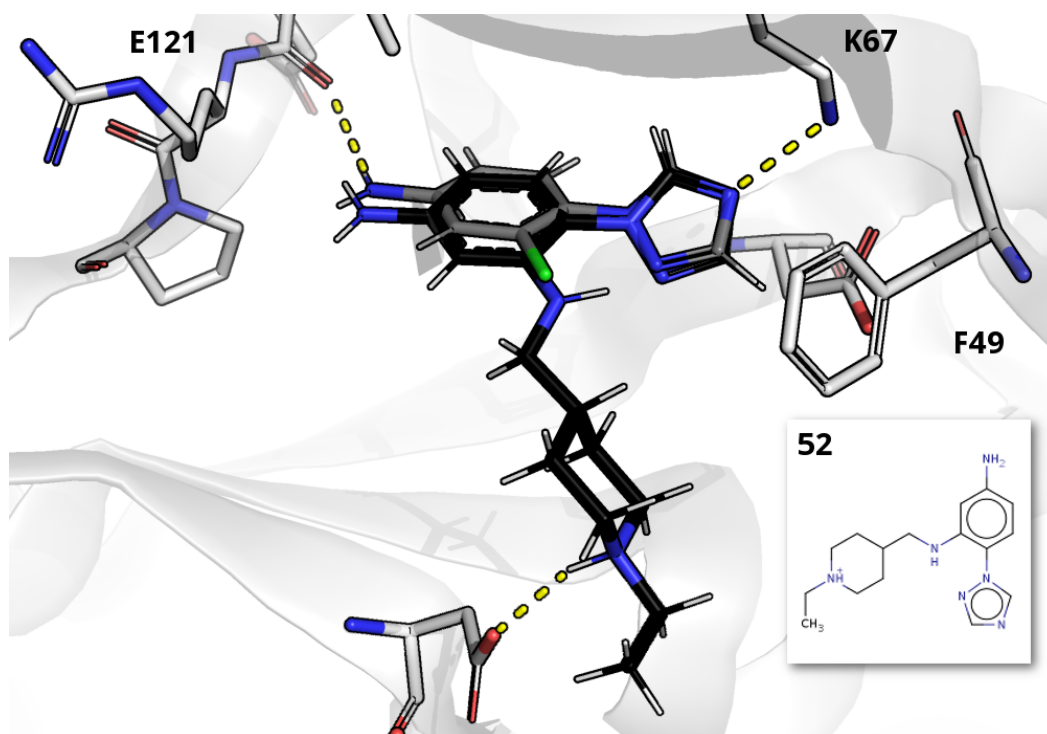


Figure 3.14: Predicted binding mode of the most potent ligand from the aniline triazole series, **52** (black carbons, stick representation), the predicted binding mode of the aniline triazole scaffold (dark grey carbons, stick representation) in the Pim-1 binding pocket, key residues (light grey, stick representation) labelled and 2D structure (inset) of the pose displayed. Predicted polar interactions are represented by yellow dashed lines. The aniline triazole scaffold (SCUBIDOO fragment 4012414) is overlayed to exemplify the fragment-based growing approach used.

None of the compounds tested from either series showed activity against PKA.

A complete list of molecules **50** to **55** can be found in the Appendix Table 7.10.

3.3.3 Project 3: novel inhibitors from the ZINC dataset

Several constraints were applied to the ZINC datasets downloaded for docking; molecules had to be purchasable, their molecular weight between 250 and 350Da, $\log P \leq 3.5$ and contain protomers within a pH range of 6-8. Applying these restrictions led to a database size of 4601 296.

The entire dataset was docked and after checks for availability, torsional strain and likely protonation states, 28 were submitted for group evaluation. Of these, 6 were selected for purchase after checking for similarity with known actives. Molecules chosen for purchase were found to be generally dissimilar to 8834 known actives against Pim-1 (T_c : $\bar{x} = 0.40$, $\sigma_{\bar{x}} = 0.09$).

As Pim-1 has a more apolar hinge, the canonical Pim-1 backbone hinge interaction tends to be large apolar moieties and this series were no exception with many molecules retrieved and 4/6 purchased having an indole predicted to be in this position (Figure 3.15). Most compounds were also predicted to interact with the catalytic lysine (K67) with some (3/6) predicted to interact with both the hinge and K67. A diverse array of chemotypes were predicted to interact with K67 but generally they were rigid, fused-ring systems (e.g. isoquinoline, triazolo-thiadiazolyl/pyrimidine).

A complete list and 2D structures of molecules **56** to **61** can be found in the Appendix Table 3.3.

A TSA was conducted on the all purchased compounds and from these, 5 molecules displayed a positive temperature shift of at least 4°C, again indicative of stabilisation of the formed protein-ligand complex. Summary TSA data can be found in Appendix Table 3.3.

Despite one compound displaying a negative shift on the TSA, all compounds were selected for crystallography. All compounds were co-crystallised [168] with the aid of a consensus peptide, Pimtide [169]. Datasets were subsequently obtained for all compounds but only crystal structures of 2 of these molecules in complex with Pim-1 could be fully refined. Protein-ligand X-ray structures were determined of molecules **57** and **61** to a resolution of 1.97 Å and 2.21 Å respectively.

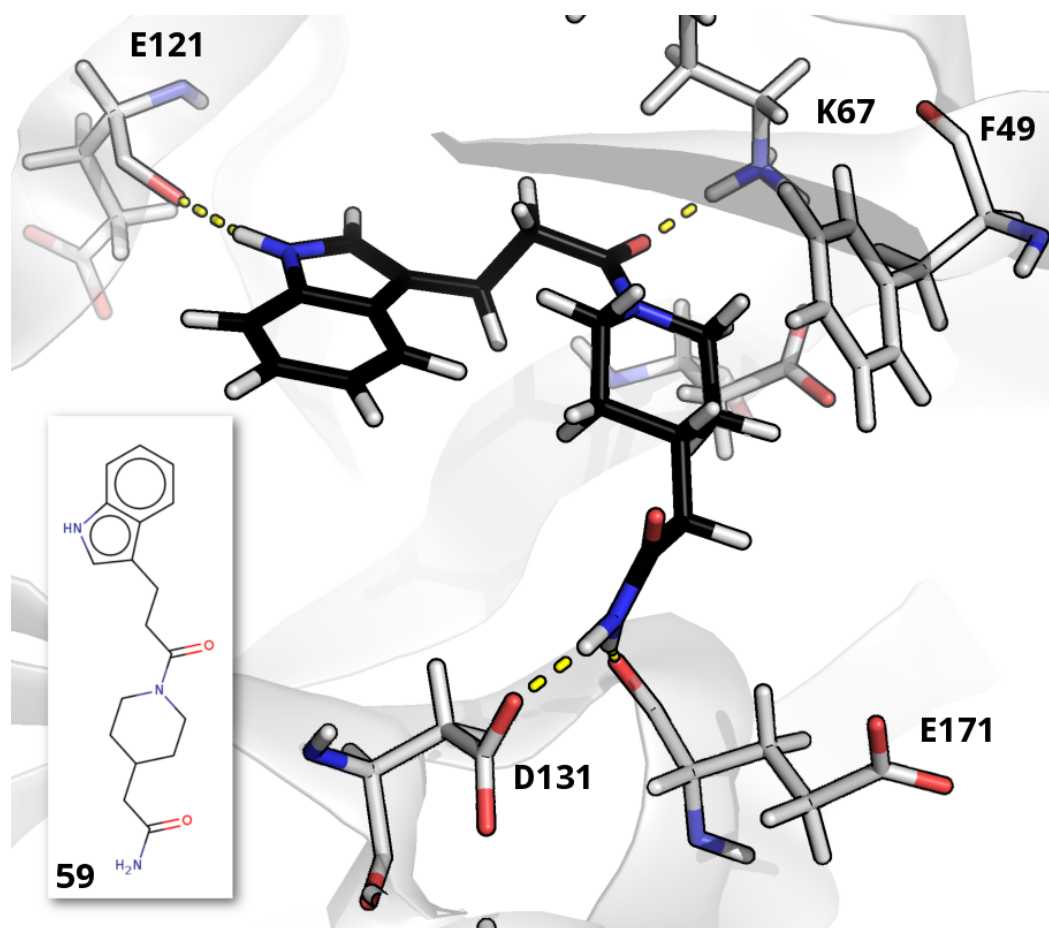


Figure 3.15: Predicted binding mode of **59** (black carbons, stick representation) in the Pim-1 binding pocket, key residues (light grey, stick representation) labelled and 2D structure (inset) of the pose displayed. Predicted polar interactions are represented by yellow dashed lines. Depicted here are interactions between the secondary amide of **59** and K67, an indole nitrogen with hinge residue E121 and its primary amide and D128.

Unexpectedly, **61** displayed quite a different binding mode in its crystal structure when compared to the docking prediction (Figure 3.16). Its isoquinoline motif was not found to be interacting with K67 as predicted. Instead it was positioned on the hinge side of the binding pocket. The secondary amide of **61** formed a bridging interaction with W278, R6 of Pimtide and D131 at the Specificity Surface and its pyrrolidine was positioned to form an interaction with D186 rather than D128 and D131 at the Specificity Surface of Pim-1.

The exact binding mode of **57** was not clear as the ligand was only partially visible in the X-ray structure. However the position of the ligand was redolent of a different binding mode than that predicted during the initial screen (Figure 3.17).

As with **61**, it appeared likely that **57** was involved in a water-bridged interaction with Pimtide. It is possible that the crystallised ligand was the hydrolysis product of the ester linking **57**'s indole and pyrrole moieties. It was, however, subsequently determined that by including Pimtide and crystallographic waters in receptor preparation, the binding

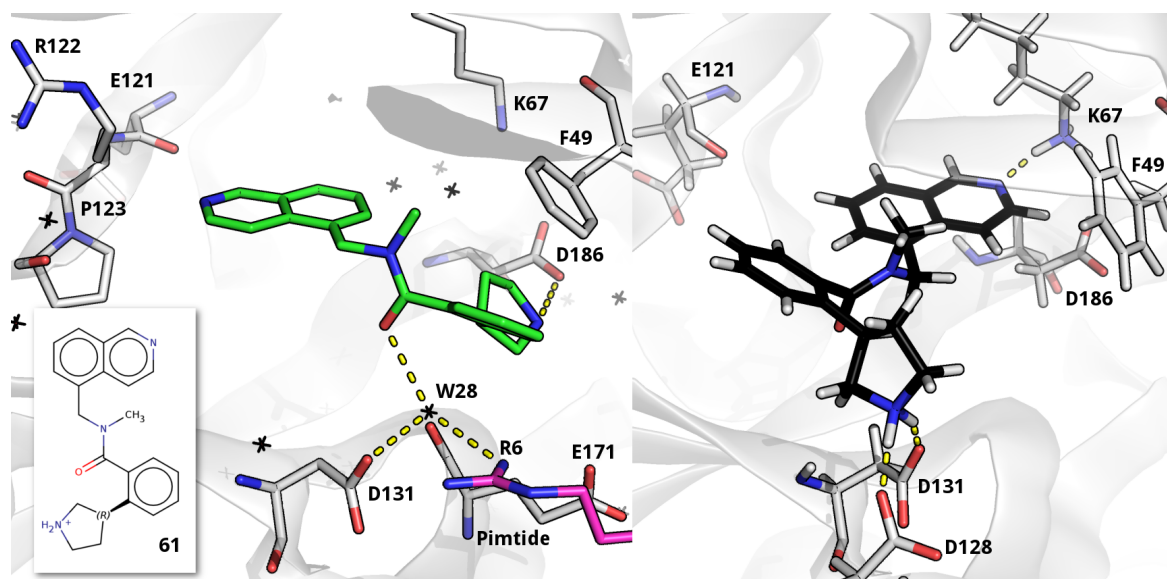


Figure 3.16: X-ray structure of **61** (left, green carbons, stick representation) in complex with Pim-1 and Pimtide (magenta carbons, stick representation) as compared to the docking prediction (right, black carbons, stick representation) with no Pimtide. Key residues (light grey, stick representation) are labelled and the 2D structure (inset) of the pose displayed. Predicted polar interactions are represented by yellow dashed lines. This depicts the very different binding mode with respect to the docking prediction when the secondary amide of **61** formed a water-mediated interaction with R6 of Pimtide and a salt-bridge between its pyrrolidine and D186. This is as compared to the predicted interactions between its isoquinoline with K67 and salt-bridge between its pyrrolidine and D128. Despite the sequence similarity of Pimtide to the known Pim-1 co-substrate, this molecule was not found to be active on orthogonal assays where Pimtide was not present and is thus unlikely to reflect an inhibitory binding mode with the Pim-1 co-substrate.

mode of the complete structures of both **57** and **61** could be replicated by re-docking to their X-ray structures (Figure 3.18).

All 6 compounds were tested against a panel of 10 kinases by a third-party vendor and their residual activities calculated. Compounds **57**, **58** and **59** showed inhibitory activity against Pim-1 specifically with **58** also reducing the activity of Pim-2 and mTOR.

Molecules were tested via FRET-based assay and K_i 's of **57**, **58** and **61** subsequently elucidated. Compound **58** was found to be a mild ligand against Pim-1 ($K_i=15.09\mu\text{M}$ with **57** showing signs of inhibition at higher concentrations (Figure 3.19). All others showed little or no activity.

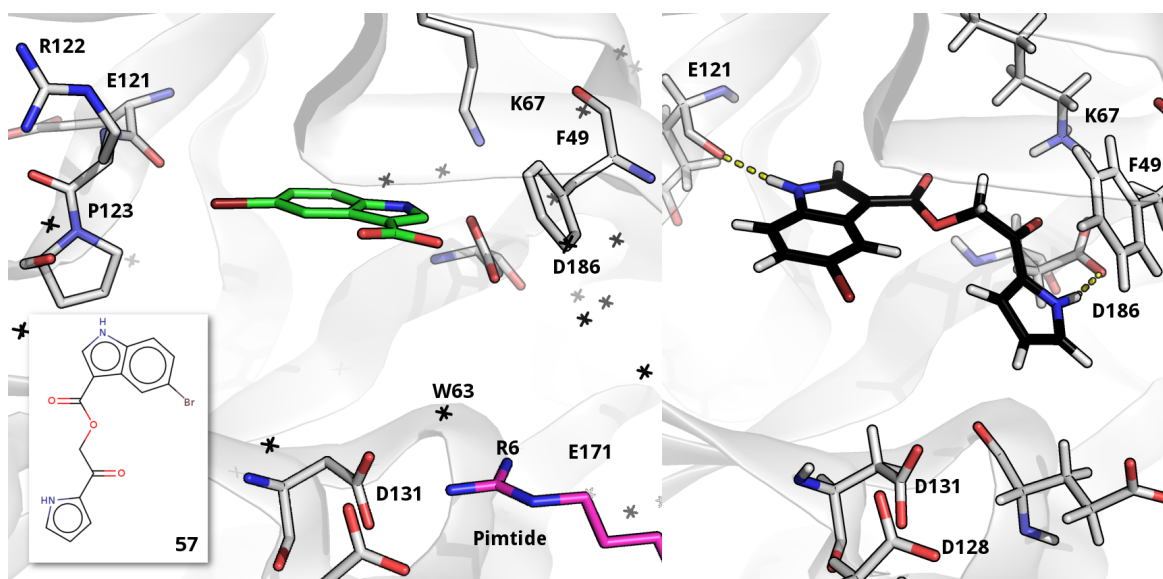


Figure 3.17: X-ray structure of **57** (left, green carbons, stick representation) in complex with Pim-1 and Pimtide (magenta carbons, stick representation) as compared to the docking prediction (right, black carbons, stick representation) with no Pimtide. Key residues (light grey, stick representation) are labelled and the 2D structure (inset) of the pose displayed. Predicted polar interactions are represented by yellow dashed lines. Although only part of the ligand was placed in the binding pocket, the difference in the binding mode can be seen as the interactions predicted by docking, substituted indole with hinge residue E121 and pyrrolidine with D186, likely did not form in the X-ray structure.

Compound **58** was predicted by FRED to interact with the hinge with an indole positioned there (Figure 3.20). An additional polar interaction was predicted, a triazolo-thiadiazole moiety predicted to interact with the catalytic lysine (K67). It also displayed the highest positive temperature shift on the TSA ($5.16^{\circ}\text{C} \pm 0.23$), indicative of higher protein-ligand complex stability.

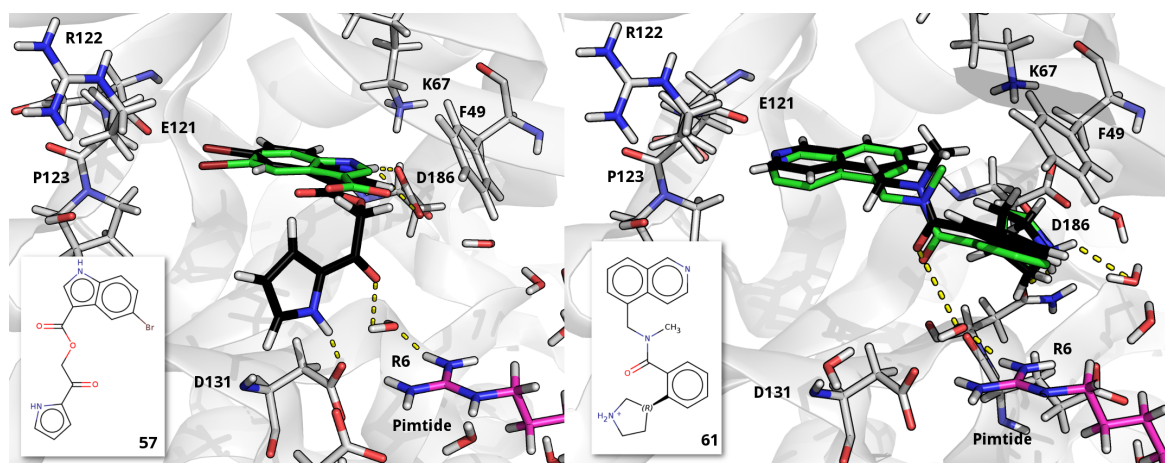


Figure 3.18: Depicted are the X-ray structure binding modes of ligands **57** (left, green carbons, stick representation) and **61** (right, green carbons, stick representation) in complex with Pim-1 and Pimtide (magenta carbons, stick representation) as compared to the docking prediction (black carbons, stick representation). Key residues (light grey, stick representation) are labelled and the 2D structure (inset) of the pose displayed. Predicted polar interactions are represented by yellow dashed lines. In both cases, only by including in the docking calculation both Pimtide and crystallographic water molecules (red oxygens, light grey hydrogens, stick representation) that were interacting with Pimtide in the X-ray structure was the X-ray binding mode of the ligand able to be recreated.

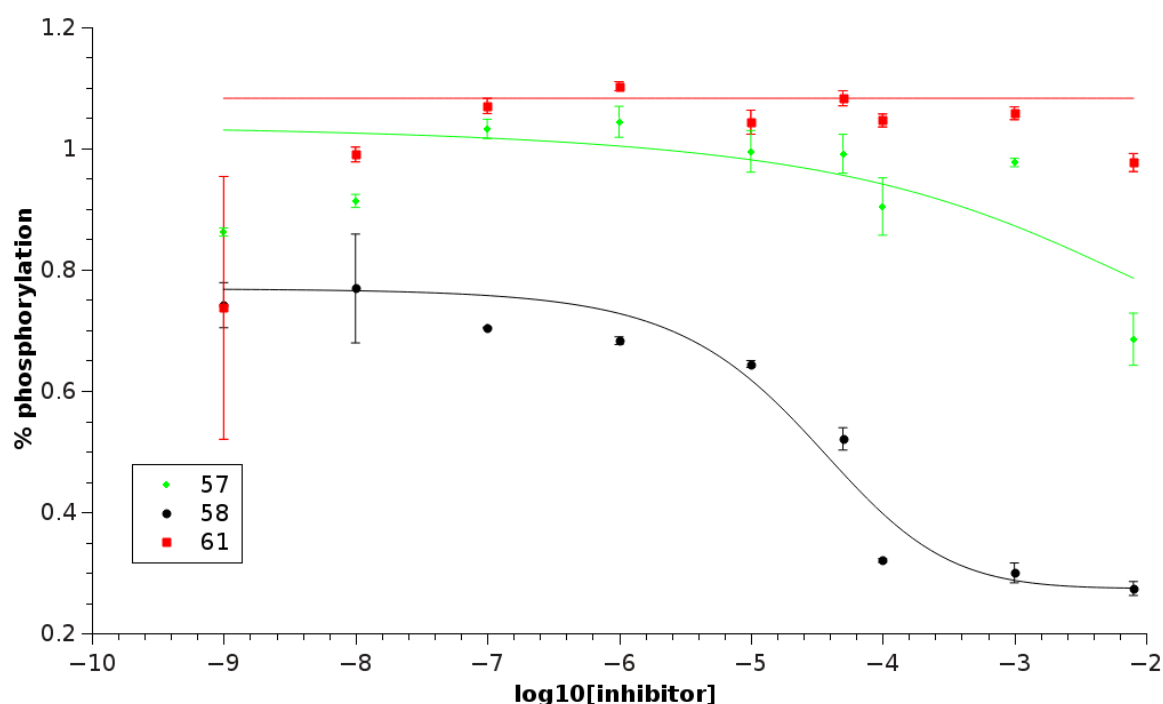


Figure 3.19: Dose-response curves of FRET-based assays of compounds **57**, **58** and **61**. Depicted here is the per-cent phosphorylation against the LOG₁₀ ligand concentration of tested compounds displaying activity against Pim-1 with error bars (standard-error of measurement) and curves fitted to the data (multi-coloured lines). The IC₅₀ of a given compound was calculated at the mid-point of the curve, indicating half-maximal inhibition of the protein. In this case, **57** (green line) is displaying mild inhibitory activity whereas **61** (red line) shows almost no activity and some indication of fluorescence at higher ligand concentrations. Staurosporine (black line) is again depicted for comparison. Each experiment was repeated with two mechanical replicates and three biological replicates

Number	ZINCID	TSA shift °C ($\bar{x} \pm \sigma_{\bar{x}}$)	Eurofins IC ₅₀ (μM)	K _i (μM)	2D structure
56	ZINC0880252	-0.5 ± 0.62	NOT TESTED	NOT TESTED	
57	ZINC09314085	4.83 ± 0.00	56.5	308.02	
58	ZINC13689545	5.16 ± 0.23	4.2	13.2	
59	ZINC40542640	2.33 ± 0.40	> 10mM	NOT TESTED	
60	ZINC71840985	2.50 ± 0.47	NOT TESTED	NOT TESTED	
61	ZINC72154357	4.00 ± 0.23	> 10mM	> 10mM	

Table 3.3: Predicted Pim-1 inhibitors, sample mean and standard deviation of temperature shifts in thermal-shift assays conducted, IC₅₀ as determined by third-party vendor, measured K_i's from FRET-based assays conducted internally and 2D structures of all ligands. The most potent ligand in orthogonal assays (**58**) also displayed the highest temperature shift on the TSA. Compound **61**, despite being crystallised in complex with Pim-1 and Pimide, showed little activity on orthogonal assays.

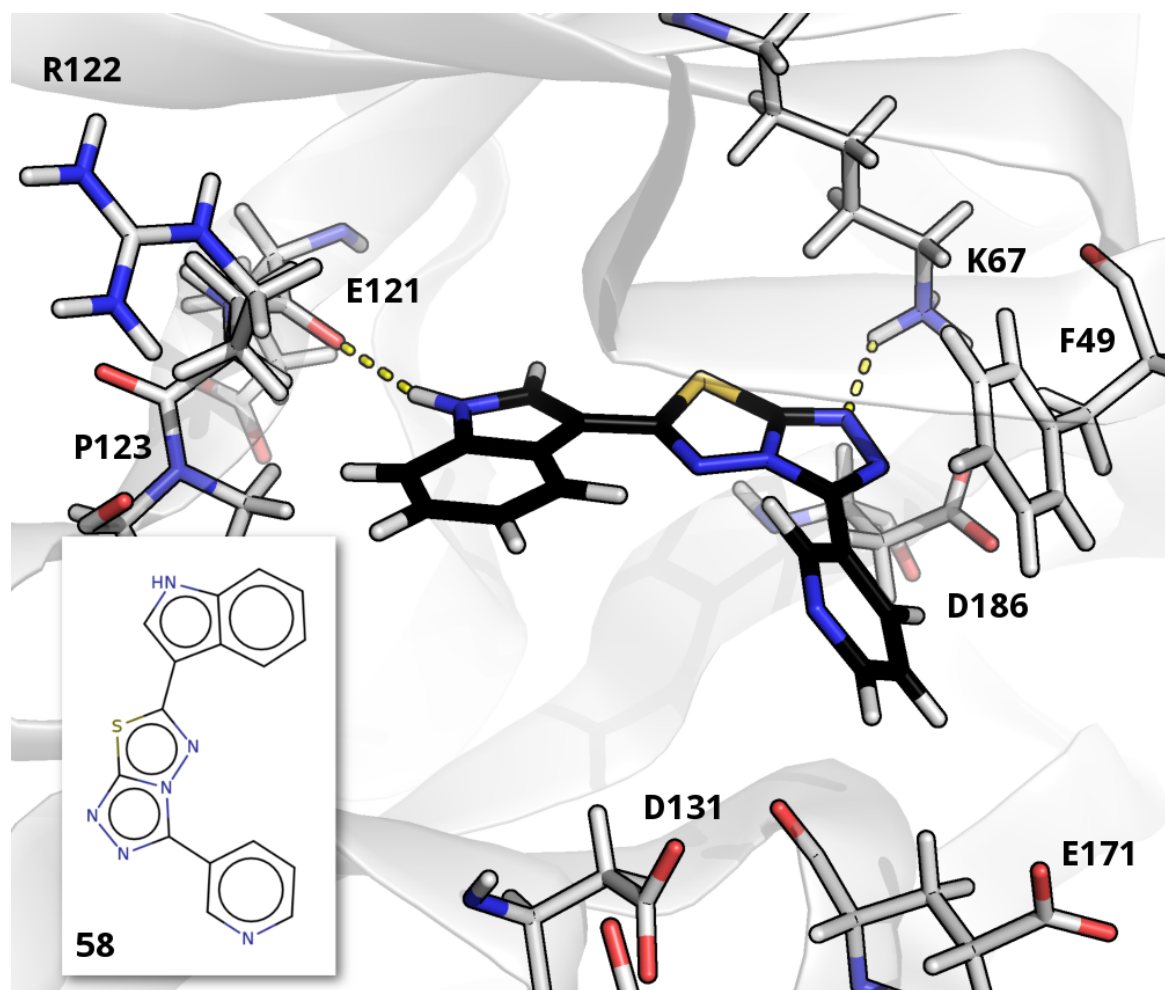


Figure 3.20: Predicted binding mode of **58** (black carbons, stick representation) in the Pim-1 binding pocket, key residues (light grey, stick representation) labelled and 2D structure (inset) of the pose displayed. Predicted polar interactions are represented by yellow dashed lines. Depicted here are interactions between the triazolo-thiadiazole core of **58** with K67 and its indole with hinge residue E121. After also being tested against a panel of other kinases, **58** was found to be active against Pim-2. As both Pim-1 and Pim-2 share significant sequence homology with Pim-3, **58** is thus a candidate selective pan-PIM inhibitor.

Activities against all kinases tested and calculated K_i 's against Pim-1 can be found in the Appendix Tables 7.11 and 7.12.

3.3.4 Project 4: extensions of truncated fragment f200

Prior to commencing any fragment elaboration it was necessary to establish the binding mode of the fragment in the binding pocket as this had implications for its stability and that of extensions to it. As hydrogen atoms are not well resolved by X-ray diffraction studies, the crystallographic structure of f200 does not provide explicit information regarding the protonation state of the ligand.

In the X-ray structure, three water molecules (W63, W62 and W29) were resolved and

arranged in a network within the binding pocket, bridging interactions with I104 and K67. One of the crystallographic waters, W62, is closest (~ 3.8 Å) to the amide of the ligand's quinoxaline core, albeit slightly beyond hydrogen-bonding distance and with apparently sub-optimal geometry for an interaction. Nonetheless, prior work [170] has suggested that W62 is the only buried water in Pim-1's binding pocket.

Calculated water hot spots from MDMix simulations of Pim-1's apo structure depict areas of elevated solvent density surrounding the positions of all three waters and much of the area of elevated solvent density adjacent to a position similar to W62 was within hydrogen-bonding range of the position of the quinoxaline core (Figure 3.21). There was significantly less solvent density surrounding a similar position of W63 suggesting waters positioned here may be displaceable to allow room for a buried water to form a polar bond with the ligand.

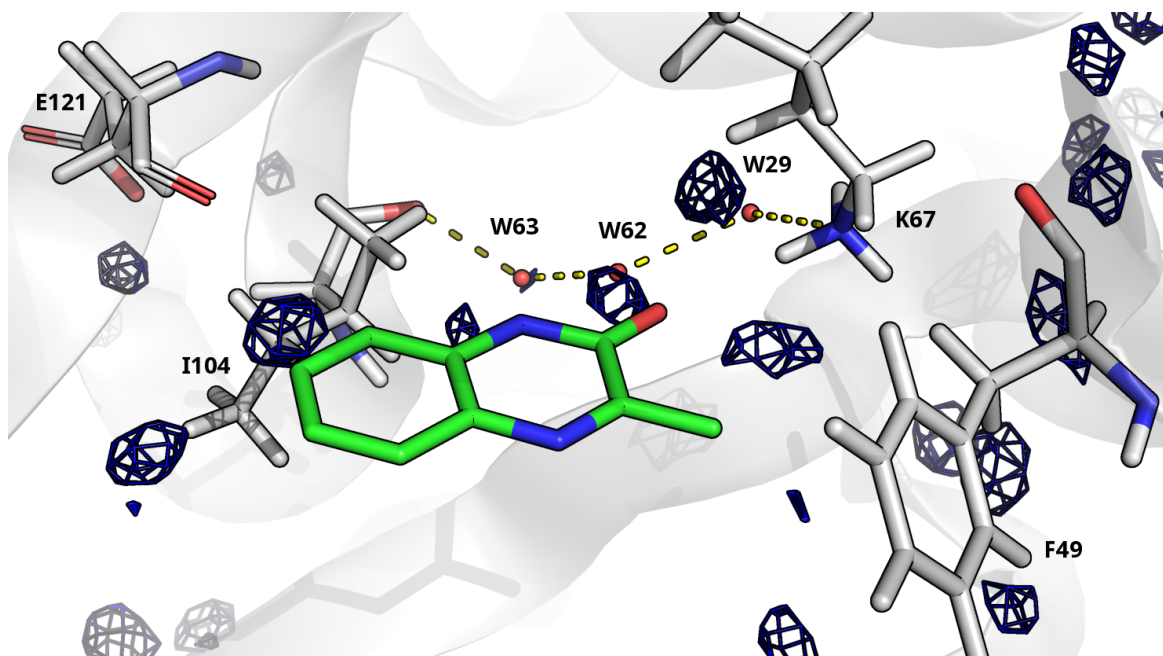


Figure 3.21: Output from MDMix showing areas of solvent density below -1.0 kcal·mol $^{-1}$ (blue mesh) overlayed with the f200 crystallographic ligand (green carbons, stick representation) and crystallographic water molecules (red spheres) to illustrate the proximity of these water molecules to calculated solvent density from MD simulations.

Further simulations of the ligand confirmed that the water in a similar position to W63 (W3) is less stable in the binding pocket, leaving it ~ 2 ns into the production simulation. This allowed an inward rearrangement of I104's side chain into the binding pocket and a water (W2) to move into hydrogen-bonding range of the quinoxaline amide nitrogen. Although the water network was disturbed by this event, the interaction between the quinoxaline amide and W2 ($\bar{x}_{DISTANCE} = 3.00$ Å, $\bar{x}_{ANGLE} = 148.46^\circ$) and another between the amide carbonyl and K67 ($\bar{x}_{DISTANCE} = 2.94$ Å, $\bar{x}_{ANGLE} = 154.33^\circ$) meant the ligand's binding mode was

stable for > 10 ns. After this point, polar interactions were broken and the ligand began to leave the binding pocket.

By contrast, simulation of a negative control, the imino tautomer of the quinoxaline's amide group, showed a close contact between the hydroxy group of the iminol and K67 after heating and equilibration when positioned similarly to the X-ray ligand (Figure 3.22). Soon after commencing production, W3 again left the binding pocket (~2 ns) and the water network was again disturbed. The ligand was ejected from the binding pocket soon after (~3 ns).

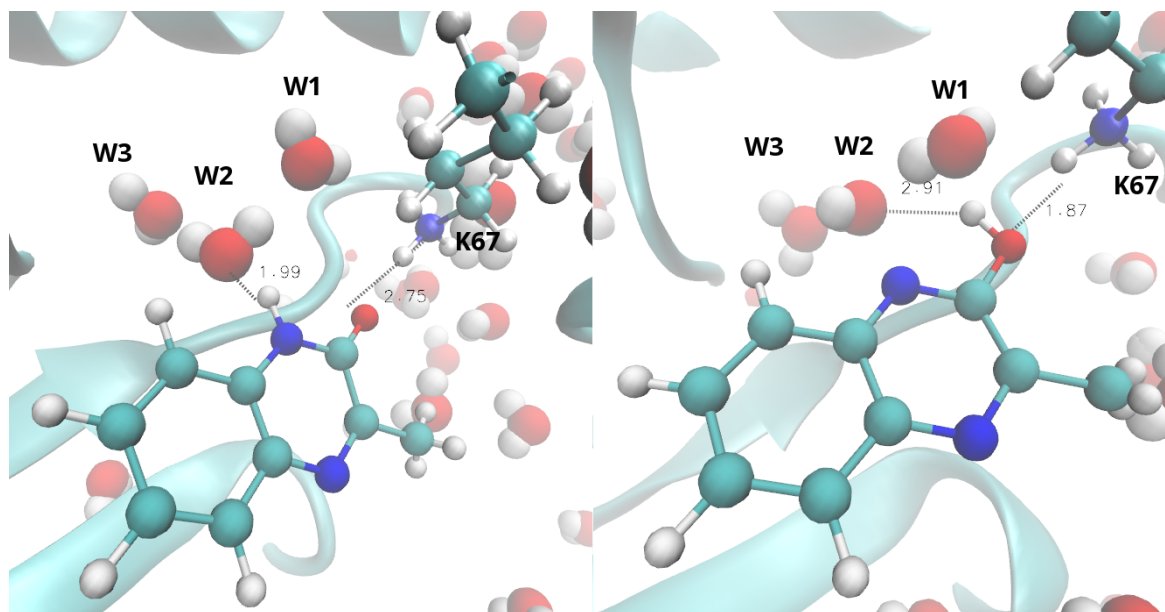


Figure 3.22: Positions and distances (Å) from an MD snapshot between ligand (turquoise carbons, ball-and-stick representation), water molecules in the Pim-1 binding pocket W1, W2 and W3 (red oxygen, white hydrogens, vdW representation) and K67 for f200 amide (left) and iminol (right) tautomers. In the former case, interactions are predicted to form between the quinoxaline's amide carbonyl and K67 as well as the amide nitrogen and a buried water molecule (W2). In contrast, the proximity of the iminol hydroxy to the hydrogens of K67 leads to repulsion and lower stability in the binding pocket as demonstrated by its shorter simulation time before interactions are broken (~2 ns) as compared to > 10 ns for the amide tautomer.

Plots of all distances and RMSDs measured during simulations can be found in Appendix Figures 7.6 and 7.7.

Reacting the ligand *in silico* resulted in a docking library of 841 molecules. The binding modes of all docked poses were again assessed by an initial visual inspection and analysis. From these an initial list of 38 molecules was selected. They were again submitted for evaluation and grading by lab members, grades were collated and analysed and 8 molecules were subsequently selected for synthesis.

As with Project 2, all molecules chosen displayed predicted interactions with residues at the Specificity Surface, specifically D128. However, two molecules displayed an alternative

predicted binding mode for the initial fragment. For **62** and **68**, the interaction between the carbonyl group of the quinoxaline's amide nitrogen is preserved but the amine is instead predicted to interact with D186 whilst retaining the interaction with D128 (Figure 3.23).

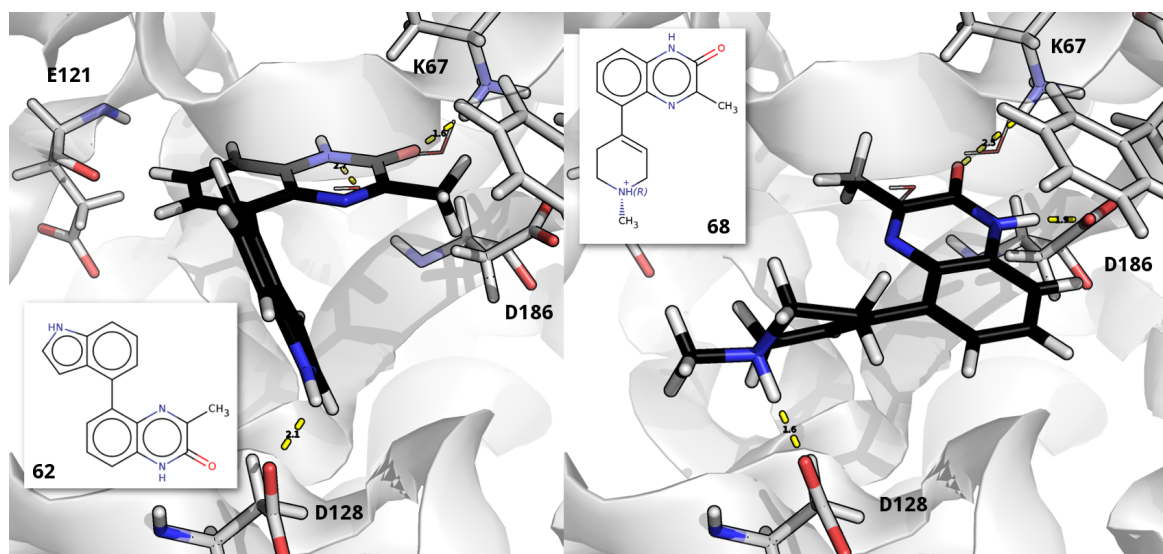


Figure 3.23: Example alternative predicted binding modes in the Pim-1 binding pocket of **62** (left, black carbons, stick representation) and **68** (right, black carbons, stick representation) in the Pim-1 binding pocket. Key residues (light grey, stick representation) are labelled and 2D structure (inset) of the pose displayed. Predicted polar interactions are represented by yellow dashed lines. In the case of **62**, the molecule was predicted to retain the X-ray crystallographic binding mode of the f200 quinoxaline and an additional polar interaction between its indole moiety and D128. An alternative predicted binding mode of **68** depicts interactions between the quinoxaline, K67 and D186 as well as a salt-bridge between its piperidine D128.

A complete list and 2D structures of molecules **56** to **63** can be found in Appendix Table 7.13.

Molecules **63**, **66**, **67** and **69** were tested and did not show activity against Pim-1.

3.4 Discussion and conclusions

That any given docking program is able to find molecular poses close to their determined crystal position has been established as an important performance criterion by prior work [171] in so-called ‘cognate’ docking. It was demonstrated that the molecular pose closest in terms of RMSD to the crystal structure is not necessarily the highest scoring pose. As we often do not know the crystal pose *a priori*, the more ‘real world’ test is to dock multiple low-energy conformations of a given molecule to the target.

As demonstrated, generation of multi-conformer molecules for docking is crucial to ensure adequate sampling of conformational space and accurate placement of the molecular pose

prior to scoring. However, the number of conformations needed to ensure adequate sampling is not straightforward. What we see from this study is that from only a few conformers per molecule ($\bar{x} = 4.5$, $\min = 1$, $\max = 24$), reliable fragment binding modes from docking were the result.

The two conformer generation and docking pipelines used in this work further support the influence of molecular shape in docking. Docking with the programs FRED, HYBRID and SEED all used the same database. Although SEED was overall superior in terms of RMSD with respect to the crystal structure as reference, FRED's performance in retrieving a given pose close to the crystal structure was acceptable considering only the top scored pose was retained. Despite generally being within the top 5% of docked poses, at no stage was the top scored SEED pose also the closest in terms of RMSD.

DOCK generally failed to recreate crystal binding modes when retaining only the top scored pose, despite manipulation of matching spheres into positions ideal for polar contacts. This is suggestive that for physics-based methods such as SEED and DOCK, one would generally need to analyse the placement of multiple poses rather than relying on the scoring function as sole arbiter of pose quality.

More influential is the conformational shape of the protein. The structure used for most docking here, 3BGP, broadly represents the most frequent conformations of ligand-bound Pim-1 structures uploaded to the PDB. This is particularly the case with respect to important flexible residues such as the catalytic lysine (K67) and residues in the P-Loop. It clearly outperformed the apo_MD structure when a given ligand had a carboxylate anion forming a salt bridge with K67. In this case, the 3BGP structure often retrieved poses $<2 \text{ \AA}$ from the X-ray position whereas the apo_MD structure failed to recreate the binding position of these molecules in all cases.

The reason becomes more clear when one compares the positions of K67 and P-Loop residues of the 3BGP and apo_MD structures. Although the average RMSDs of residues themselves did not deviate much during the course of the apo_MD simulation, the positions of both are very different with respect to the two residues in the X-ray structure (Figure 3.24). Thus, the shape complementarity of the binding pocket was disturbed. The next effect was that the apo_MD structure did not retrieve poses close to that of the fragment in the X-ray structure, even when using software (HYBRID) where information regarding the likely binding site was supplied by the position of the ligand from the geometry in 3BGP.

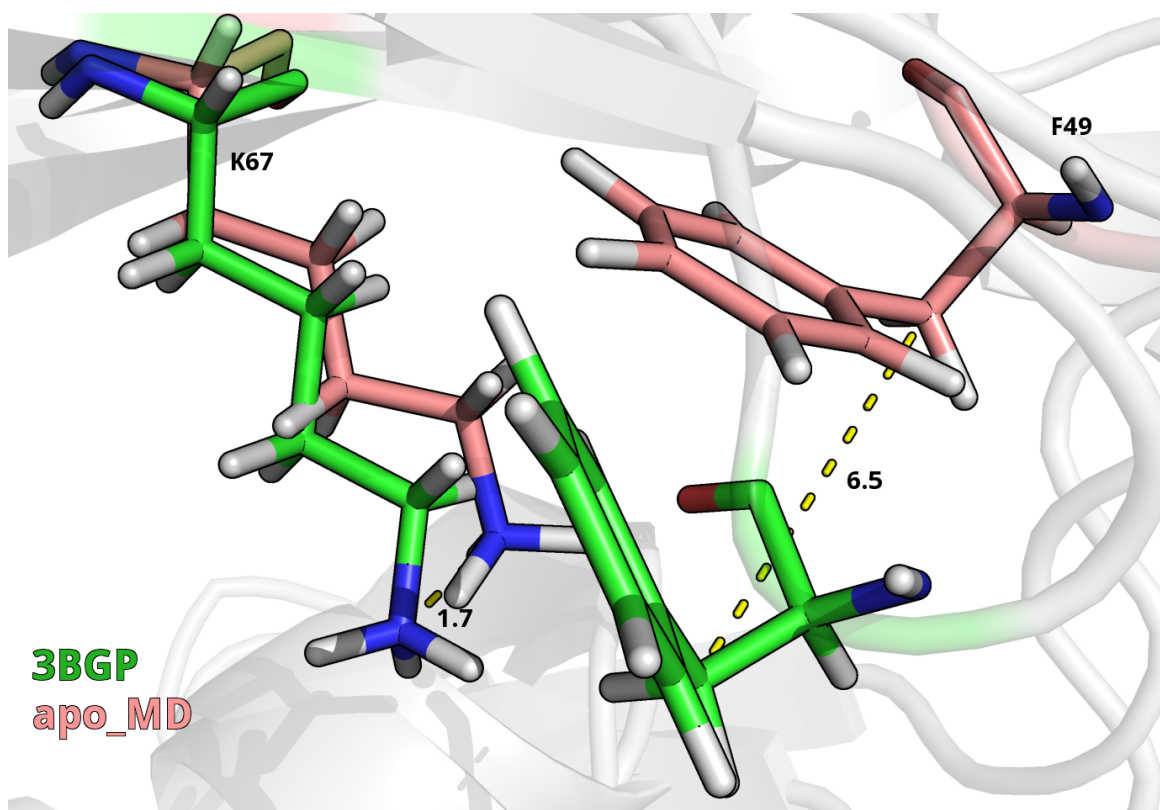


Figure 3.24: Depicted are differences between the positions of K67 and F49 (P-Loop) in 3BGP (green carbons, stick representation) and apo_MD (pink carbons, stick representation) structures. This is indicative that in the absence of a bound ligand, both F49 and K67 retract into the binding pocket, negatively affecting docking codes' ability to recreate X-ray binding modes as shape complementarity is disturbed.

RMSDs and number of poses retrieved by HYBRID, SEED and DOCK were improved or remained similar when docking to a snapshot from MD trajectory taken from the simulation based on the 3BGP X-ray structure. This is suggestive that simulating with a bound ligand will result in a conformation that reflects the ligand's binding mode but one that is also suitable for the retrieval of new ligands. However, results were significantly better with HYBRID as opposed to FRED so some binding site information is still required.

Alternatively, this also suggests that with knowledge of the binding pocket, preferring HYBRID over FRED with a simulated structure reduces the probability of prioritising an unlikely fragment binding mode for further development. This lends further support to protein conformation being the most important criterion when deciding upon a starting structure for rigid docking.

From this, a workflow can be expected (Figure 3.25). Where binding site information is augmented by a bound ligand, superior docking results were obtained by simulation of a bound ligand structure followed by minimisation and then docking with software that integrates binding pocket information e.g. HYBRID. This approach is benefited by reducing the

reliance upon the researcher's choice of starting structure.

Although HYBRID outperformed both FRED and DOCK generally, DOCK outperformed HYBRID in some cases, such as when an interaction with the catalytic Lysine (K67) was predicted to interact with a moiety on the ligand weaker than a salt-bridge. Salt-bridges are generally over-estimated by force-field methods [172] but it may be that HYBRID, with its focus on shape, underestimates the importance of ligands possessing weaker interactions. Whether this occurs in HYBRID's scoring step, where interaction potentials are smoothed, or in its Chemical Gaussian Overlay (CGO) step, where the docked pose is directly compared to the bound ligand, is unclear.

In the absence of binding pocket information, it remains that the choice of starting protein structure generally requires expert intervention or analysis of multiple protein structures. Following this, good results can be obtained via a more traditional work flow of unbiased fragment docking to establish the most likely binding mode by analysis of multiple poses followed by a second round of docking with programs optimised for larger molecules e.g. FRED.

Fluorescence assay results from fragment extensions in Project 2 were somewhat mixed but go some way to confirming that compounds targeting residues at the Specificity Surface of Pim-1 require a more rigid structure. Previous ligands had either large rigid rings or rigid amide linkers and **26**, the most potent ligand, was the only compound tested that had an unsaturated linker between a benzoic acid moiety and a substituted benztetrahydrofuran predicted to directly interact with D128. Similarly, **52**, although possessing a saturated linker between its aniline triazole moiety and its piperidine, the molecule was likely made more rigid by an internal hydrogen-bond between its triazole and a secondary amine in the chain.

The specific moiety interacting with Specificity Surface residues is also likely affected by their proximity to the bulk water boundary, an area where relatively low-permittivity of a protein's binding pocket meets the relatively high permittivity of bulk water. D128 is quite close to where the boundary likely is and prior work [173] has elucidated that solvent polarisation effects at the protein-solvent interface have a 'dampening' effect on energies and forces acting upon pair-wise electrostatic interactions. More specifically, interactions between charged atoms are particularly decreased by this effect (solvent screening) [174].

Although the estimation of overall interaction energy for all tested molecules was predicted to be favourable by docking and after minimisation, in every case where a charged moiety

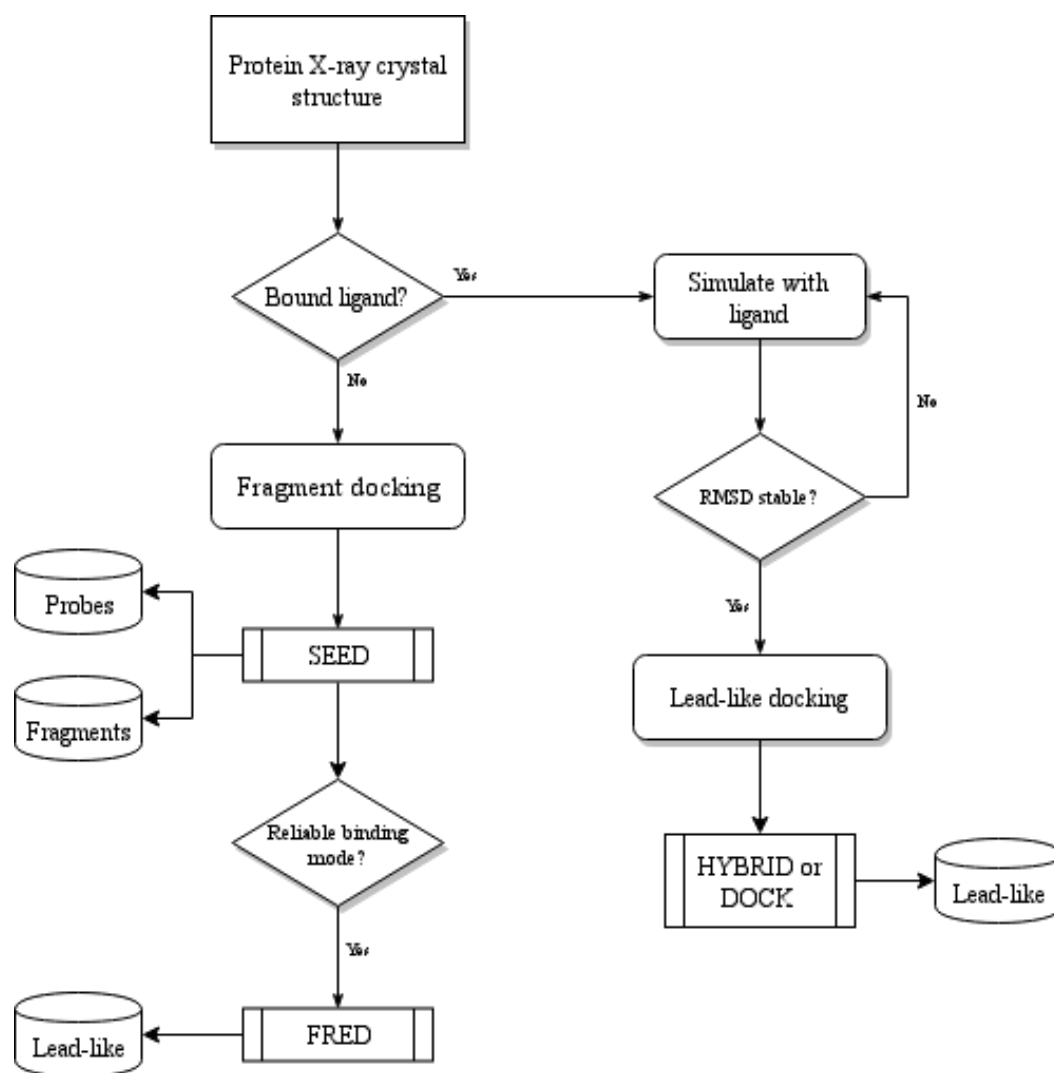


Figure 3.25: Suggested workflow for various docking software when coupled with molecular dynamics simulation. Recommended tasks are modelled by rounded rectangles, decision points by diamonds, recommended software by squared rectangles and databases by ovoid cylinders.

was predicted to interact with D128, solvent screening effects were predicted to be high. Thus, molecules with moieties such as a piperidine or charged primary amine in this position displayed either poor or negligible affinity in the assay. Other compounds, such as **52**, had a piperidine predicted to interact with D128. But the piperidine nitrogen also has an ethyl substituent, possibly negating solvent polarisation effects. However, despite possessing a measurable affinity for Pim-1, it was a relatively poor ligand inhibiting in the three-digit μM range.

In any case, charged molecules forming salt bridges in a protein binding pocket are traditionally prioritised highly in docking. Where residues are within the binding pocket, these are likely to be very strong interactions. This is borne out by the number of ATP-competitive inhibitors for Pim-1 where a salt bridge is formed with the charged side-chain of K67, in-

cluding many of the inhibitors in this work. However, although not a definitive indicator of the likelihood of an interaction, the results of this study make clear that a reduction in the relative strength of interactions with solvent-exposed charged residues should be considered when analysing a given molecule's overall binding characteristics.

The trend with regard to inhibitors with predicted charged interactions was confirmed in Project 3. Any inhibitors in possession of charged moieties predicted to interact with D128 (**60** and **61**) showed little to no activity against Pim-1. The most potent compound from this series, a triazolo-thiadiazole (**58**), was predicted to bind in ATP-competitive fashion and did not address D128 at all. Although an X-ray structure was not obtained of this ligand, it achieved the highest measured positive temperature shift on the TSA. It inhibited at the two-digit μM range or better on orthogonal assays and appeared to be a selectively pan-PIM inhibitor. Its potency and position within the binding pocket therefore offer a reasonable starting point for a molecular optimisation campaign.

Two inhibitors were crystallised and their X-ray structures solved, **57** and **61**, the latter the first X-ray structure on the PDB to display a direct interaction with Pim-1's consensus peptide (Pimtide). The binding modes of both compounds were only able to be recreated by re-docking them to their crystal structures with Pimtide and its crystallographic waters in place in the prepared structure. Compound **61**'s pyrrolidine was found to be interacting with D186 in the DFG loop, the only X-ray structure found to be interacting with this residue without also interacting with the charged side chain of K67.

Additionally, whilst **61** showed little activity in the absence of Pimtide in our FRET assay, **57** was mildly active in this assay, an orthogonal inhibition assay and showed a positive temperature shift in the TSA performed in addition to being crystallised with Pimtide. Structurally, **57** possesses a substituted indole and pyrrole acceptor. These are frequently reported amongst known Pim-1 inhibitors and their typical placement is predicted by the docking binding mode i.e. the indole placed at the hinge interacting with E121 and the pyrrole predicted to interact with K67. This is suggestive of multiple binding modes for this ligand but this was not confirmed.

Simulations of the quinoxaline fragment and a negative control in Project 4 tentatively confirmed the stability of its binding mode within the binding pocket of Pim-1. A network of three waters between I104 and K67 is found in many Pim-1 X-ray structures but simulations here have demonstrated that, in line with previous work, only one of them (W29), bridging with K67, is likely to be conserved in Pim-1 generally. Simulations have also demonstrated

that upon the ejection of the least stable water, closest to I104, there is a rearrangement of the remaining waters. One of them, W62, formed a direct interaction with the protonated amide nitrogen of the ligand's quinoxaline core. A new network of interactions was thus formed involving direct interactions with the quinoxaline's carbonyl and K67 in addition to the those already described (Figure 3.22).

The X-ray binding mode of the quinoxaline fragment was somewhat corroborated by simulations but, unfortunately, activity of the fragment itself was not detected in FRET-based assays. This is presumably due to the low affinity of fragment sized-molecules, a common phenomena in FBDD campaigns [43]. Four synthesised extensions were also tested and found to be inactive against Pim-1.

The exact reason for lack of activity by these compounds in the FRET assay is unclear. However, all of the predicted inhibitors were extended *in silico* from position 6 of the quinoxaline core. Whilst this allowed for extensions that were predicted to interact with D128, none of the apolar sections of extensions tested were particularly close to the hinge.

By way of contrast, in addition to possessing greater rigidity and a predicted interaction between D128 and an uncharged moiety (a hydroxyl), the apolar benzofuran sidechain of **26** was also predicted to be more proximal to the more apolar Pim-1 hinge. It therefore seems reasonable to suppose that inhibitors extended *in silico* at position 6, rather than 5, may more effectively address the Pim-1 hinge in docking studies and possess the added benefit of greater shape complementarity in the Pim-1 binding pocket (Figure 3.26).

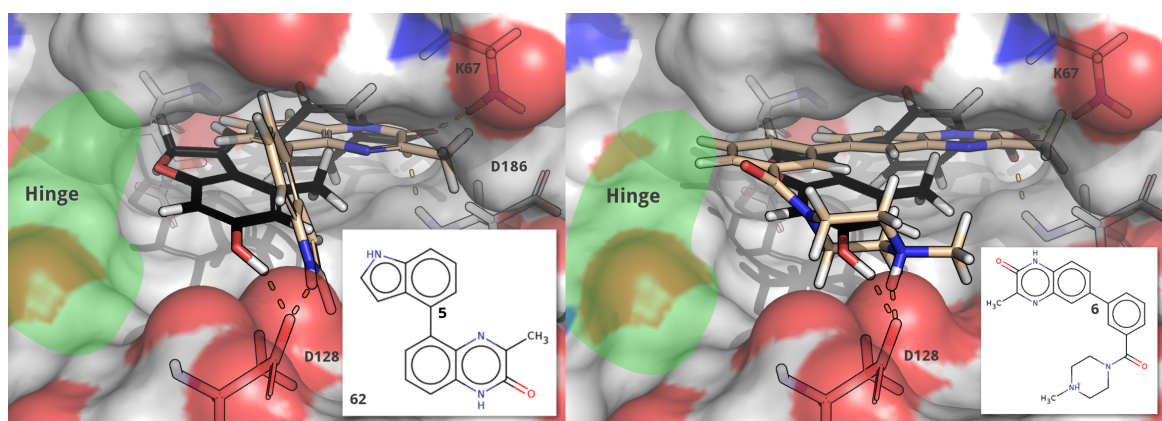


Figure 3.26: Predicted binding modes of **26** (black carbons, stick representation) vs **62** (left, straw coloured carbons, stick representation) and an example docked pose (right, straw coloured carbons, stick representation). Depicted here is a prediction of decreased distance and enhanced shape complementarity to the more apolar Pim-1 hinge region (green region) after Suzuki Coupling *in silico* at position 6 to the quinoxaline core from Project 4 as opposed to extending from position 5.

It bears asking the question whether completing analysis, synthesis and testing in-house, as per Project 2 and Project 4, is a more efficient approach than purchasing ligands and using 3rd-party vendors for testing as per Project 3. Broadly comparing the two approaches, using a large dataset such as ZINC, curated to be lead-like, is the more reliable method purely for finding new ligands. PINGUI is pitched as an ideas generator and on that front it succeeds as it certainly did generate many possibilities for synthesis and testing in Project 2 and Project 4.

However, its utility is dependent upon the presentation of a reliable fragment binding mode. This is, as demonstrated, difficult to achieve without expert knowledge in choosing the target protein structure. Related, upon determining a reliable fragment binding mode, it was found that only few molecules actually retained it upon extensions being added.

Assuming the docking program gets the binding mode right, more molecular possibilities are always useful. But the claim of synthetic feasibility is also questionable as, in almost all cases, the molecule tested was either synthesised using a different method due to the risk of low yields or synthetic complexity or for similar reasons, a derivative with a minor change (e.g. pyridine \rightarrow aryl ring) was proposed. This necessitated additional work to confirm the binding mode of the derivative prior to any attempts at synthesis as less than half of the molecules were synthesised with the reaction as proposed.

There was also a vast difference in both the computational and personnel effort to develop a similar number of inhibitors at a similar range of potency against Pim-1. This appears to be the case even when the focus was on synthesising specialist inhibitors to address specific areas in the Pim-1 binding pocket. Using the PINGUI-developed library in Project 2 necessitated compiling and parametrising an initial library for docking and separate docking runs on a HPC cluster to avoid results being dominated by *in silico* reactions that naturally tend to generate a lot more products than others. This was followed by further filtering steps prior to analysis. All of this was followed by synthesis effort that was not straightforward and subsequent affinity testing. Compare and contrast this to merely downloading and docking a ZINC subset as in Project 3, an order of magnitude larger than the dataset in Project 2, followed by affinity testing and structural information gained from X-ray crystallography.

However, the costs of completing all steps in-house must be balanced against the gains. From Project 1, structural information about the Pim-1 binding pocket was gained as decisions regarding starting X-ray structures and which software to use in subsequent virtual screening campaigns were simplified.

In Project 2, more was learned regarding the likelihood of successful novel ligand synthesis and how minor changes to ensure synthetic feasibility can also ensure subsequent ligands. A chemotypically broader array of molecules synthesised and tested elucidated structural information about Pim-1 itself as we now know several moieties that are both less likely to interact with charged residues near the bulk-water interface but also how to ameliorate these solvent polarisation effects.

Finally, more was learned regarding the stability of Pim-1's structural waters from Project 4 and how this leads to a stable binding mode for a novel scaffold for Pim-1. From a stable fragment binding mode, synthetically feasible products were the result from using only one reaction at one site on the quinoxaline's core structure. More sites are available to be used to explore the Pim-1 binding pocket.

What constitutes a 'hit' is also an important question posed by this work. A hit, of course, requires experimental confirmation. However, there are degrees to this. Through a significant amount of work, a few inhibitors in the μM range or worse were the result. So perhaps one may not consider these relatively weak ligands real hits. These compounds are, however, not optimised for binding to Pim-1 and the two clear hits from synthesis projects, **26** and **52**, are chemically diverse and provide excellent starting points for optimisation campaigns.

There were essentially three hits from Project 3. Compound **61** may not be considered a real hit as it only crystallised via an interaction with Pimtide and therefore showed no assay activity. This is despite the likely similarity between Pimtide and the substrate in cellular conditions. Its activity in the absence of Pimtide could not be assessed as Pimtide's sequence (ARKRRRHPSGPPTA) fits within the peptide sequence in the assay kit used. However, **57** and **58** were both active in orthogonal assays. Despite not progressing to full refinement in crystallographic studies, **58** in particular appears to be a selective pan-PIM inhibitor, albeit at the μM level.

Additionally, the binding properties of Pimtide with respect to ligands demands further investigation. The water-mediated interaction of **61** with Pimtide involved a charged moiety and occurred near the bulk-solvent boundary (Figure 3.27). Using Pimtide potentially obviates the need for a direct interaction between a charged ligand and charged residues that are difficult to target for similar reasons to those near the Specificity Surface (D128-D131). These residues in addition to those near the Ribose pocket (E171) and the Phosphate Groove (K169, D167) offer the possibility of novel and selective inhibitors for Pim-1 as there are few published inhibitors targeting them and they are less conserved in kinases.

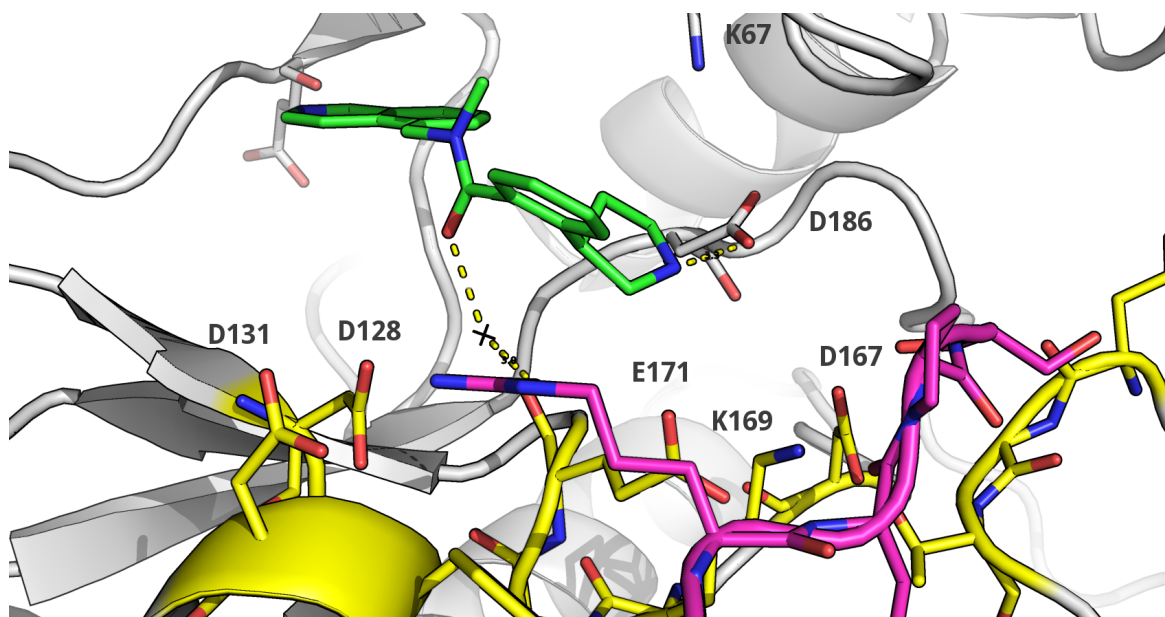


Figure 3.27: Crystal structure of **61** (green carbons, stick representation) in complex with Pim-1 a water-mediated interaction (red, cross representation) and Pimtide (magenta carbons, stick representation). Key residues (yellow, stick representation) are labelled and predicted polar interactions are represented by yellow dashed lines. Depicted here are charged residues near the Ribose pocket and Phosphate Groove (see Figure 3.3) that could potentially be targeted in the Pim-1 binding pocket for the development of peptide inhibitors.

Pimtide's binding position and proximity to the binding pocket also offer the possibilities for the design of peptide inhibitors. It was reported in a review of computational and lab-based studies that peptide inhibitors can be viable strategy for the development of substrate-competitive inhibitors, particularly for Ser-Thr kinases [175]. Computational approaches are complicated by the conformational flexibility of peptides and *in vivo* work is made more difficult by their poor stability and bioavailability. However, Pimtide's rigidity and the predictability of its binding mode with Pim-1 suggests a role as a Protein-Protein Inhibitor, of which there has been recent success using structure-guided methods with more flexible peptides against a Tyrosine kinase, ABL [176]. This is, however, quite separate from the ATP-competitive inhibitors discussed in this study.

So what makes a good ligand for Pim-1? As intimated by the results presented here, rigidity is important for high-affinity binding to Pim-1, particularly on the hinge side of the pocket. With respect to the more apolar Pim-1 hinge, a very recent optimisation campaign [177] confirmed that rigid, often large ring systems where vdW interactions dominate, are important for positioning the ligand in the absence of the second conserved hydrogen-bond donor usually found in the hinge of kinases. The ligands from that study, a series of tetrahydropyrrolo quinazolinones, were made more rigid and planar by the presence of an intramolecular hydrogen-bond between the pyrrole-NH to the quinazolinone-N and made large potency

gains on this basis (Figure 3.28, A). Indeed, the necessity of planarity was made clear by an study [178] of activity cliffs with respect to a series of benzofuranones, where a series of hinge-binding pyrrolo pyridines were synthesised where the only difference was the position of the pyridine Nitrogen. Even a slight disturbance of planarity resulted in large decreases in affinity.

The search for novelty may therefore be more fruitful from the side of the Pim-1 binding pocket occupied by its catalytic Lysine, K67. Another very recent optimisation campaign of diamino pyrazoles [179] reported pM inhibition whilst targeting residues near the Specificity Surface (e.g. D128, E171) (Figure 3.28, B). Their results suggest that the necessity for rigidity on the hinge sided of the binding pocket is perhaps less important on the Lysine side, offering some flexibility in terms of shape-complementarity and, consequently, in what moieties can target less conserved or solvent-exposed residues in this region. The binding position of ligands such as Staurosporine, ostensibly a hinge binder, does *prima facie* contradict this. Its very large fused aromatic ring system, however, does also satisfy the need for rigidity. Nevertheless, for future campaigns targeting more novel residues, strategies involving chemotypes extending from the catalytic Lysine side of the Pim-1 pocket would appear to be more effective at finding more affine and potentially selective ligands.

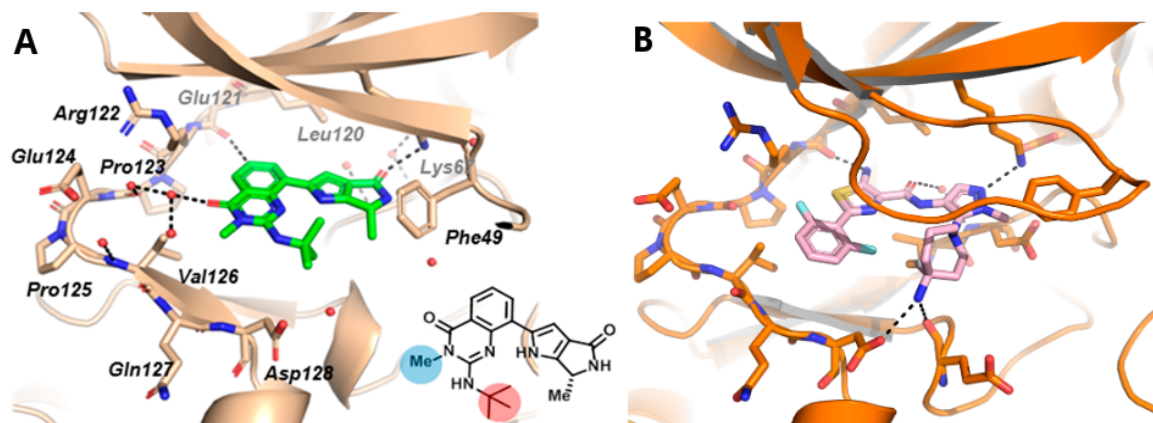


Figure 3.28: Example ligands suggesting future targeting strategies for Pim-1. (A) Depicted is a ligand (green carbons, stick representation) with multiple rigid endocyclic systems in the Pim-1 binding pocket with crystallographic water molecules (red, spheres). Key residues (tan carbons, stick representation) are labelled and likely polar bonds are represented by dashed black lines. The ligand also forms an intramolecular hydrogen-bond between its pyrrole-NH and the quinazolinone-N that preserves the planarity of the ligand (adapted from Wang, H-L, et al 2019) (B) In this case, a ligand (pink carbons, stick representation) in complex with Pim-1 (orange, cartoon representation) targets solvent-exposed residues (e.g. D128) by extending from the catalytic Lysine side of the Pim-1 binding pocket, a strategy that resulted in ligands with improved shape complementarity (adapted from Wang, X, et al 2019)

To summarise, whilst Project 3 was a more rapid method to find novel ligands based in existing knowledge of Pim-1 and kinases generally, more was learned from the other projects

that will guide subsequent ligand optimisation campaigns and generate productive streams of research in and of themselves. On balance, the net gain in terms of chemical novelty and structural understanding of Pim-1 and kinases is therefore larger with the latter. Promising results from structural studies should also inform the development of Pim-1 peptide inhibitors, further adding to structural information about Pim-1 generally. The data gained from all of these projects should guide structure-based optimisation of Pim-1 inhibitors toward and beyond sub-nM affinity.

Chapter 4

Novel ligands for the Smoothened receptor

In conjunction with my supervisor, Prof. Dr. Peter Kolb, I was responsible for the overall research planning, strategy and execution of all projects. I was also responsible for completing all computational work. All assay testing of suggested compounds was carried out by collaborators at the Siebold group at Oxford University led by Prof. Dr. Christian Siebold, namely Dr. Eamon Byrne, Ms. Maria Klimopoulous, Mrs. Rebekka Siebold-Schwab and Ms. Rachel Woolley. The unpublished structure was graciously provided by the Siebold Lab.

4.1 Introduction

Members of the Hedgehog (Hh) signal pathway, first identified in studies on *Drosophila* embryonic development, play a foundational role in the cellular communication and development of vertebrate organisms. Of the large families of secreted peptide factors, for example, Wntless-related integration site (Wnt), fibroblast growth factor and TGF- β , Hh provides one of the more arresting examples of conservation of function between developmental regulation of a quite diverse array of organisms [180].

In vertebrate systems, three main genes have been identified as conserved; *Desert Hedgehog* (Dhh), *Indian Hedgehog* (Ihh) and *Sonic Hedgehog* (Shh). Expression of these genes regulate the specification of cell fate, proliferation and survival depending on the tissue being addressed. The pathway is therefore implicated in the vast majority of vertebrate tissue formation and, aside from tissue maintenance and repair, is dormant in adult cells [181] [182].

Given these important regulatory mechanisms, misregulation of the Hh pathway via various activating, inactivating and loss-of-function mutations unsurprisingly leads to an array of medical issues. Primarily, these are various cancers including medullablastomas, glioblastomas, breast cancers and one of the most common human forms, Basal Cell Carcinomas (BCCs) [183].

Within the Hh pathway, a membrane receptor that has been identified as amenable for inhibition is the Smoothened receptor (SMO). SMO is a G Protein-Coupled Receptor (GPCR) with a traditional 7-Transmembrane (7TM) bundle, four Extra-Cellular Loops (ECLs) and an extra-cellular Cysteine-Rich Domain (CRD), believed to play a crucial role in regulation of its activity [184] (Figure 4.1). Despite sharing little sequence identity with other human GPCRs [185], the designation of SMO as a GPCR has been controversial [186]. However, it has been determined that SMO exhibits the characteristics of a GPCR, demonstrating that coupling to $G\alpha$ is required for Hh signalling *in vivo* [187].

SMO's activity in the Hh pathway is largely dependent upon its trafficking with other proteins in and out of the primary cilium (Figure 4.2). Upon binding of one of the three Hh signals to a 12-Transmembrane receptor, Patched (PTC1), SMO is phosphorylated by the kinase GRK2 and free to accumulate in the primary cilium via coupling to heterotrimeric G proteins and β -Arrestin. Kinesin-like proteins such as Kif7 facilitate the formation of a special compartment at the tips of the primary cilium [188] that allow for complexation of SMO-mediated downstream signal transduction mechanisms, such as phosphorylation and

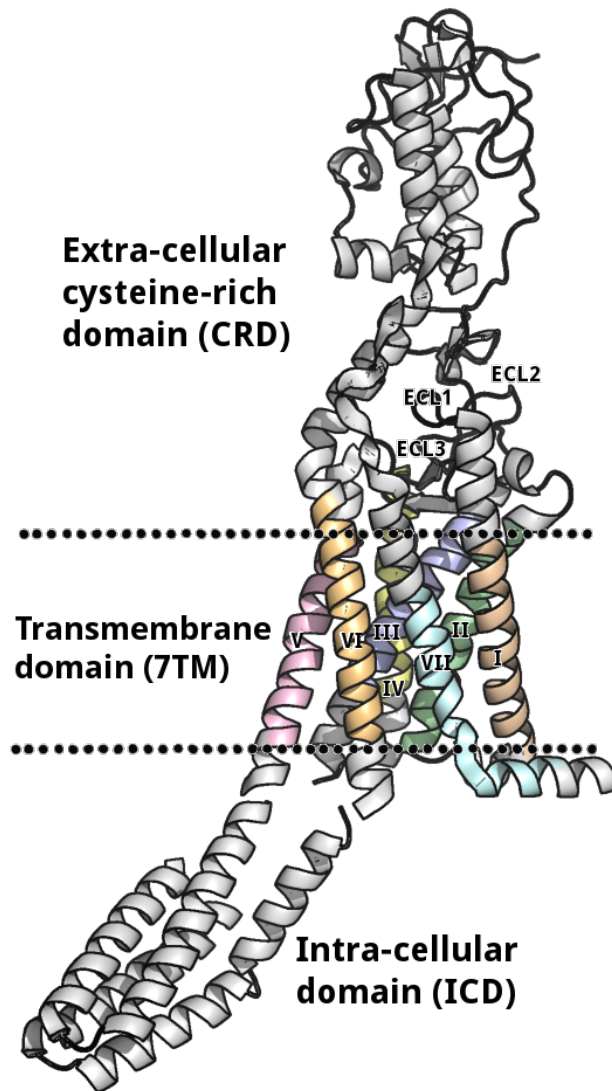


Figure 4.1: The complete SMO assembly, inactive conformation (PDB:5L7I). Depicted here are SMO's extracellular loops (CRD, grey cartoon representation), its heptahelical transmembrane domain (7TM, multi-coloured cartoon) and intracellular C-terminal domain (ICD, grey cartoon). Allosteric binding in the CRD results in conformational changes in the other domains that impel SMO toward an active conformation.

therefore activation of transcriptional effectors (e.g. Zinc finger protein Gli1). Full length Gli proteins dissociate from the unphosphorylated Kif7 and the Gli-repressor SuFu [189], bypass proteolytic processing, enter the nucleus and activate cellular transcription.

Absent a Hh signal, SMO remains bound to PTC1 and Gli proteins are instead isolated by Kif7 and SuFu to be differentially phosphorylated by PKA, GSK3 β and CK1. Gli proteins are further processed into transcriptional repressors. These truncated Gli proteins pass through the cytoplasm and into the cell nucleus to repress transcription.

The activities of SMO and PTC1 are, thus, crucial determinants of disease. At least 85% of inactivating PTC1 and 10% of activating SMO mutations drive uncontrolled replication in

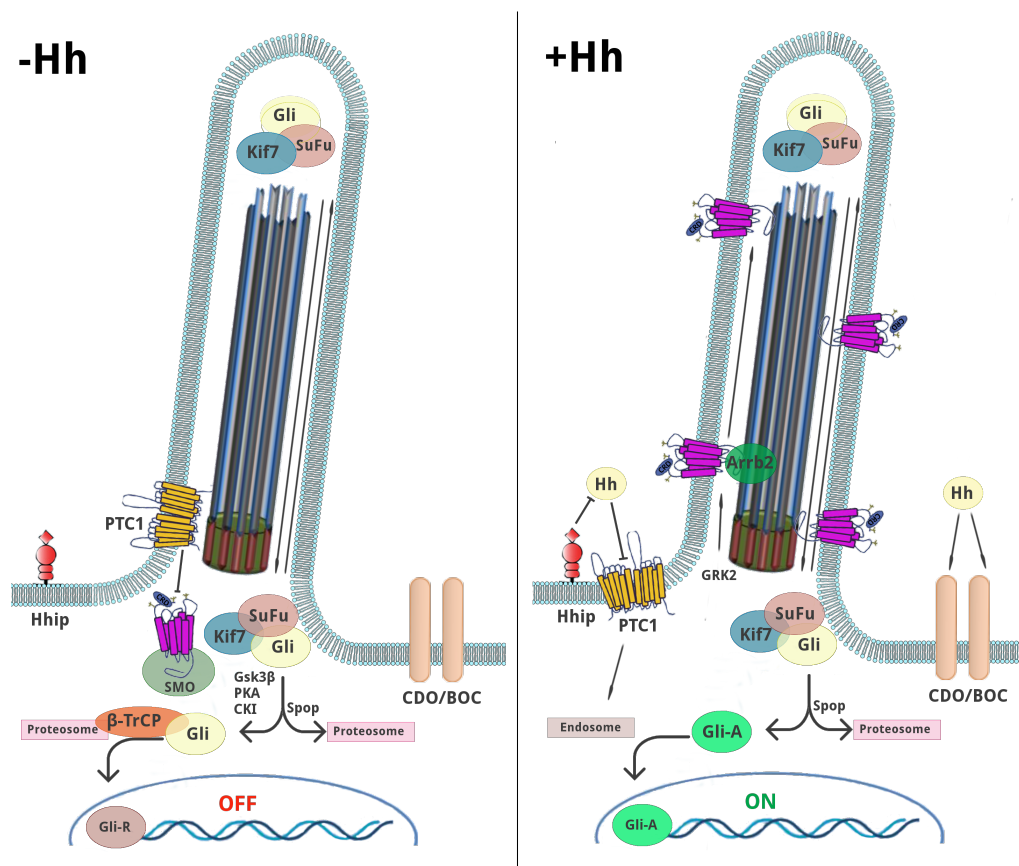


Figure 4.2: The Hedgehog (Hh) pathways in cells, activated and inactivated. Depicted here are known pathway proteins (ovoid spheres, multiple colours), signal molecules (yellow, spheres), proteasomes (red, rectangles), the endosome (light brown, rectangle), ciliary microtubule assembly (large cylinder assembly) and lipid membrane with membrane proteins CAM-related/down-regulated by oncogenes (CDO), Brother Of CDO (BOC) and Hedgehog Interacting Protein (HHIP) that bind Hh signal molecules. Depicted here, in the absence of a Hh signal, SMO is held in complex with Patched (PTC1) at the base of the primary cilia and is therefore unable to prevent Gli proteins from being cleaved into transcriptional repressors, hence, gene transcription is inactive. In the presence of a Hh signal, PTC1 is complexed with HHIP and the Hh signal and SMO is able to translocate to the tip of primary cilium where it accumulates and prevents Gli cleavage. Gli proteins are subsequently phosphorylated and activated by Kif7 and SuFu where they can pass into the nucleus and activate transcription (adapted from Teglund, et al. 2010)

Hh tumours [190]. Despite this and suggestions of other targets downstream of SMO [191], particularly in avoiding drug resistance [192], SMO is thus considered the most 'druggable' protein in the Hh pathway [193].

The first crystal structure [194] of SMO provided structural information regarding antagonism via a known small molecule inhibitor to SMO. Of other small molecule inhibitors against SMO, a particular success was the clinical drug Vismodegib (Erivedge) [195], a more polar molecule than earlier inhibitors such as Cyclopamine. However, a missense mutation in TM6 (D473) common to many BCCs has been reported to disrupt Vismodegib binding [196]. Attempts to circumvent resistance with ligands that avoid D473 has led to ligand binding unaffected by the mutation to D473 by Taladegib [197], used in trials for solid tumours

and other cancers (DrugBank Accession Number: DB12550) (Figure 4.3). Other work has prioritised novelty via an inhibitor binding deep within the 7TM bundle (SANT-1) as well as alteration of the binding pocket by an agonist, SAG1.5 [198], inhibitors that assiduously avoid D473. Although D473 is the most common mutation in BCCs, others have been reported [199]. A tool compound containing a large phthalazine core, TC114, was also used to elucidate specific structural roles played by ECL3 and TM6 for crystallographic study [200].

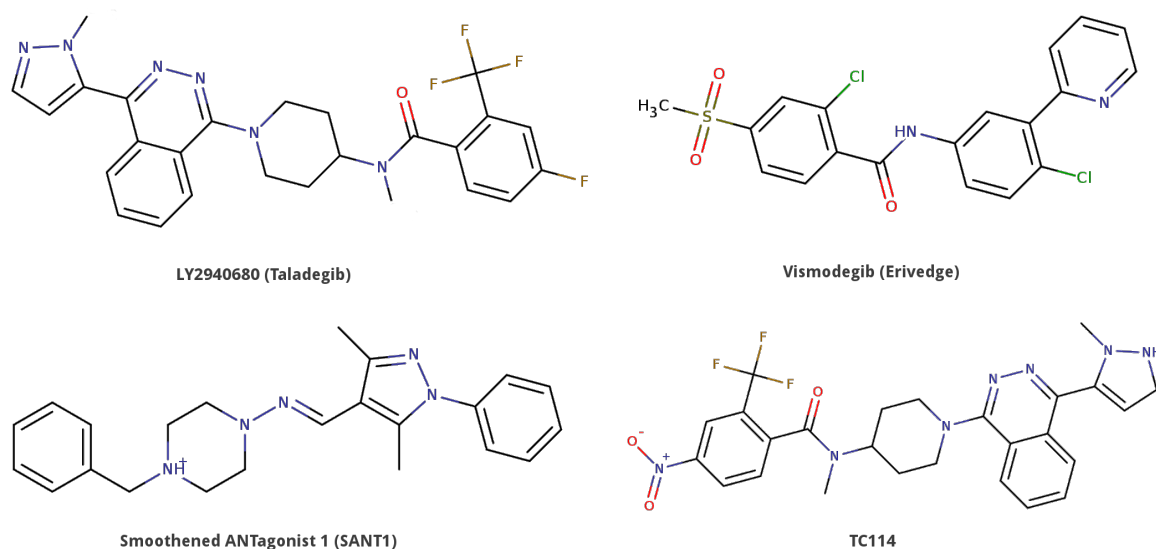


Figure 4.3: Some published drugs and inhibitors against SMO. Taladegib has previously been subject to adenocarcinoma and solid tumour Phase I and II clinical trials treating at Stages I - IIIB that have completed or it has been withdrawn from the trial. Vismodegib is currently in clinical trials for patients suffering from Basal Cell Carcinomas, pancreatic and prostate cancers at various stages and has been subject to numerous other clinical trials. TC114 and SANT-1 are not clinical candidates.

There is a sizeable gap in the literature regarding SMO inhibitors and chemical diversity. This is in terms of chemotypes, ligand binding modes and susceptibility to mutations as many SMO ligands rely on only one or two polar interactions, primarily binding via apolar interactions. Experience thus far with Vismodegib suggests it would be prudent to explore alternative binding modes, multiple positions within the TM7 bundle and more interactions to stabilise a ligand within the binding pocket.

In light of the structural information obtained from previous campaigns, Structure-Based Virtual Screening (SBVS) offers the possibility of large-scale structure-based ligand discovery for SMO. Large-scale docking campaigns have well-known limitations [9]. Nonetheless, by fitting libraries of small molecules into a binding pocket and scoring them with a function containing well-developed physics-based terms and parameters, a relatively large amount of chemical space can be quickly sampled.

Finding novel chemotypes is also more likely with large and diverse libraries when compared to a rational design campaign as in previous studies. Rapid testing of molecules can also be facilitated by using libraries comprised of commercially-available molecules as ZINC Is Not Commercial (ZINC [131]).

More recent work has elucidated structural information regarding the role of cholesterol as an agonist in the extra-cellular CRD [201] and rearrangement of various loops that allow communication across the cell membrane. Although SMO's endogenous ligand remains unknown [202], it has been found that cholesterol is both necessary and sufficient to activate SMO [203]. Binding of antagonists (e.g. Vismodegib) in SMO's Trans-Membrane Domain (TMD) induce conformational changes in ECL3 that propagate to the extra-cellular Cysteine-Rich Domain (CRD) and occlude the cholesterol binding site. This intrusion of ECL3 leads to side chain rotations of W109 and R161 and the formation of a cation- π interaction between them where cholesterol would ordinarily bind. This interaction suggests that a ligand binding in the TM7 domain should therefore prevent binding of cholesterol but the reverse is not necessarily the case.

Until recently, no ligand was crystallised in complex with SMO where cholesterol has bound to the CRD. However, an unpublished X-ray structure of active-state SMO with bound cholesterol was recently solved where a large region of difference (Fo-Fc) electron density was noted in the TM7 region, indicative of a bound ligand. A diverse array of compounds such as assay kit ingredients, PEG, protein expression compounds, or hydrolysed cholesterol were ruled out by collaborators (see Table 7.14 for a complete list).

This work therefore has several aims. Firstly, attempts to identify the unknown molecule(s) in the area of electron density (referred to as the Completely Unknown Region of Electron Density, or, 'CURED' for the rest of this work) were made with a mixture of datasets. Second, the use of small molecule probes to analyse the allosteric binding pocket of SMO and to suggest which moieties would be favourable or unfavourable in different regions. Finally, the discovery, testing and further development of novel inhibitors for SMO.

4.2 Material and methods

4.2.1 Receptor preparation

Three-dimensional crystal structures of SMO were prepared from an unpublished X-ray structure and two downloaded from the Protein Data Bank [59] (PDB: 5L7I and 4N4W). The resolution of each structure was reported as 3.2 Å, 3.3 Å and 2.8 Å, respectively.

The unpublished X-ray structure was prepared in order to identify the Fo-Fc density present within the ECL and TM7 domains of the receptor. Docked molecular probes were used both to narrow the range of possibilities of chemical moieties within sections of the CURED and 'map' the binding pocket into regions favourable for these moieties. These sections were grouped according to which moieties were favourable.

The aim of working with the 5L7I X-ray structure was to prioritise retrieval of poses from a drug-bound structure, that being Vismodegib. Vismodegib displays few polar interactions (D384), its position primarily stabilised by apolar interactions. So priority was given to docked poses that were predicted to display more polar interactions connecting ECL1 (N219), ECL3 (D384, Y394, R400) and TM6 (E518) whilst avoiding other residues in ECL3 (D473) and some in TM6 (Q477, E481) where mutations have been demonstrated to affect binding in many Hh-driven cancers [199]

An alternative structure (PDB:4N4W) was chosen to find more 'deep' binders in the region occupied by SANT-1. It has two polar interactions with residues in ECL3 (Y394) and TM6 (H470) and two apolar moieties than penetrate deeper into the binding pocket. Poses were therefore prioritised if their positions were predicted to be below the position of H470 whether they were directly interacting with H470 or not.

Missing atoms or residues for all structures were repaired and minimised with Wit!P [132], hydrogens added and their positions minimised with CHARMM v31b2 [78]. Protonation states of histidine residues in all structures were determined via visual inspection. Each histidine was allocated a status of HID, HIE and HIP for docking dependent upon whether the ϵ , δ or both nitrogens in the imidazole side-chain were considered likely to be protonated.

A complete list of protonation states for histidines in all prepared SMO structures can be found in Appendix Table 7.15

The unpublished receptor structure was prepared for docking with two program suites: Solvation Energy for Exhaustive Docking (SEED) version 3.3.5 [60] and DOCK 3.6 [68]/3.7 [24]. Receptors downloaded from the PDB (5L7I and 4N4W) were also prepared for docking with DOCK. In all cases, the X-ray ligand was removed prior to docking.

For SEED, defaults regarding departures from ideal hydrogen bonding geometry, clashes, etc. were used and residues comprising the binding site of the protein were determined via manual inspection. Pre-defined rules regarding bond-length, angle and direction of H-bond

donors and acceptors allow the construction and distribution of vectors of unitary length on all hydrogen bonding groups of ideal geometry. To reduce computations of unlikely polar-bond geometries, coordinates of known ligands were used to restrict the number of force vectors evaluated to those not exceeding a set geometry criterion ($10\text{-}70^\circ$).

The Fo-Fc density in the unpublished structure (Figure 4.4) was located primarily in extracellular regions of the SMO binding pocket but also extending deep into transmembrane regions of the receptor. The top of the electron density was located just below F484 and extended between other residues in TM6, including the hydrogen-bond network connecting it to ECL3 (D473, E518, R400), finally extending to lower residues in TM6 (H470).

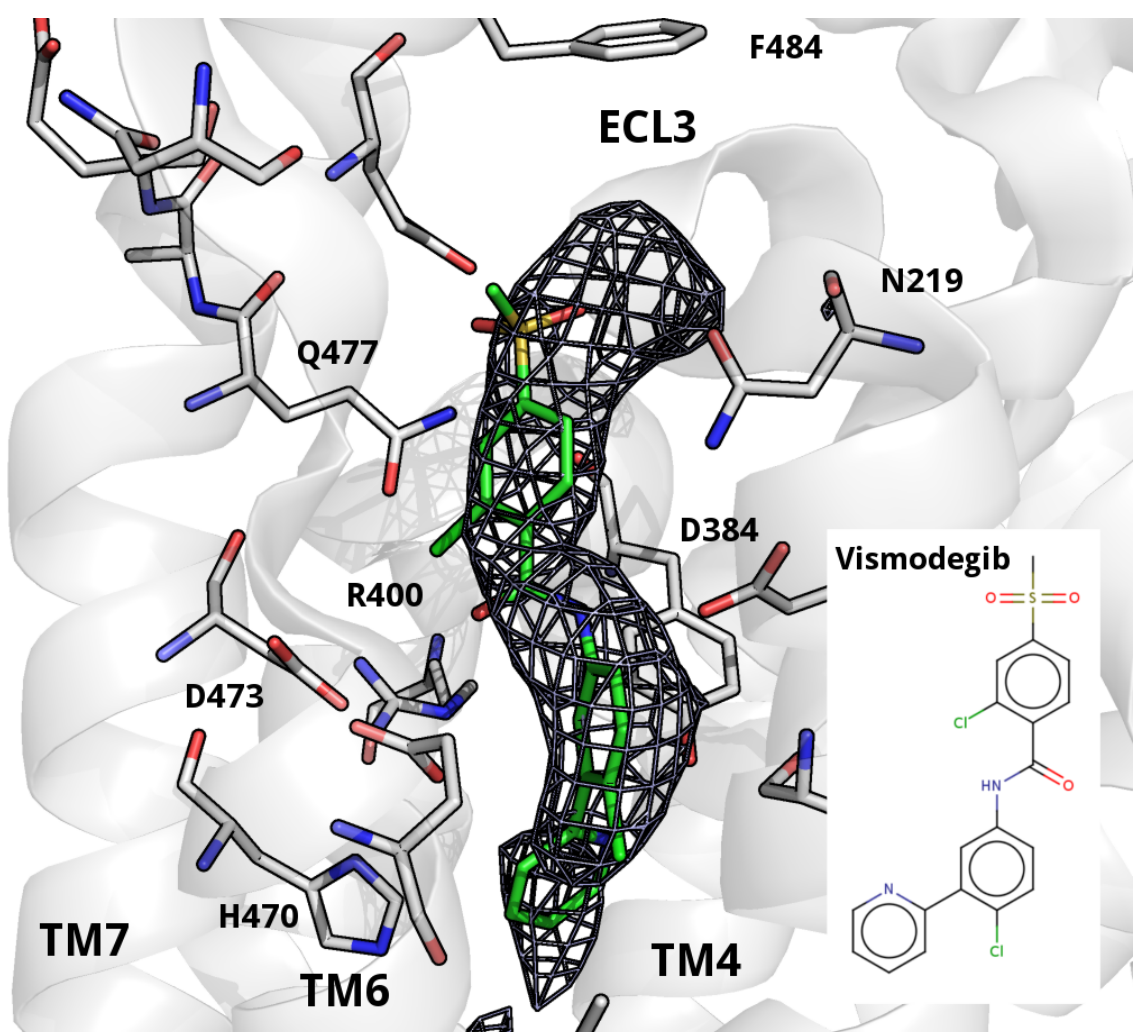


Figure 4.4: A region of Fo-Fc density (dark grey, mesh representation) from an unpublished X-ray structure is depicted here with the aligned crystal structure of Vismodegib (green carbons, stick representation) for reference. Key residues (light grey, stick representation) are labelled and 2D structure (inset) of Vismodegib is displayed. This demonstrates the binding position and orientation of the CURED with respect to the binding position of a known ligand.

In total, vectors around 29 residues were determined to provide adequate coverage of the

binding pocket around the electron density. An image of the prepared binding pocket can be found in Appendix Figure 7.8

For docking with DOCK, receptors were prepared via the generation of a set of spheres representing the invagination of the binding pocket. The geometric positions of these spheres were generated using *sphgen* [133]. The default number and positions of spheres was deemed inadequate for coverage of SMO 's binding pocket so more spheres were added and their positions set within the binding pocket to ensure coverage. This also increased the number of matching points around key residues for scoring and evaluation of poses.

After generation of vdW and electrostatic grids for the entire protein, a box was generated from the cluster centre of all spheres so that it enclosed them plus some margin. Within the boundaries of this box, scoring grids [65] describing protein energy potential functions were pre-computed for later docking and scoring. The default box size is 5 Å³ but this was generally increased.

This decision was made as, often, poses were placed partially outside of the boundaries of the default box. The effect of this was that some atoms were not considered in scoring. As the eventual score is a delicate balance involving atomic pair-wise calculations of multiple quantities, components of the overall score for affected molecules were inaccurate, as was the overall score.

Images of the prepared binding pockets can be found in Appendix Figures 7.9, 7.10 and 7.11.

A receptor was also prepared from 5L7I with an alternative conformation. All residues within 8 Å RMSD of a known ligand against SMO, Vismodegib, were identified and minimised using CHARMM and, for comparison, MOE [150].

Minimisation was completed with CHARMM using the CHARMM force field, steepest descent and conjugate gradient optimisation with Modified Partial Equalization of Orbital Electronegativity (MPEOE [204]) atomic partial charges calculated.

Minimisation with MOE was completed with the Amber10 [205]: Extended Huckel Theory (EHT) [206] force field as implemented in MOE, a method that combines Amber10 and EHT bonded parameters for large-scale energy minimization. Optimisation was completed with the Truncated Newton [207] method and atoms were parameterised with AM1-BCC [139] partial charges.

Certain residues in the 4N4W binding pocket were subject to a process known as ‘tarting’, the purpose being to increase ligand preference for specific parts of the protein binding pocket. Applying this idea, the partial charges of specific atoms in a protein are redefined and the net effect is an overall increase in the polarity of certain bonds. This increased polarisation is generally performed on backbone atoms, their partial charges increased by 0.4e prior to calculating electrostatic potential grids for docking. To encourage retrieval of poses as deep in the binding pocket as SANT-1 residues 4 Å from its position and below the position of H470 were tarted.

4.2.2 Ligand preparation

An initial benchmarking campaign was performed to assess the performance of chosen SMO structures. This campaign took a list of molecules found active against SMO from ChEMBL [12] and used these to generate a set of challenging ‘decoys’ from the DUD-E website [208]. Using 3D similarity fingerprints, DUD-E decoys are similar in terms of chemical properties (e.g. molecular weight, number rotatable bonds) to known actives but are unlikely to bind to the target. This provides a useful metric of scoring ranks vs known activity, otherwise known as *enrichment*. All known actives and generated decoys were docked to both of the 5L7I and 4N4W X-ray structures.

In order to identify the CURED, an internal library of 141 molecular ‘probes’ that includes small molecules such as benzene, water and cyclopentane was used in initial docking studies. A full list of these probes can be found in Appendix Table 7.16.

Additionally, the ZINC12 [131], ZINC15 [14], The Human Metabolome DataBase (HMDB 3.0 [209]), and the Kyoto Encyclopedia of Genes and Genomes (KEGG [210]) and a small Pharmacognosy library of natural products from the University of Vienna were used. Both ZINC12 and ZINC15 are freely-available databases of biologically-relevant and purchasable compounds with additional web-based resources such as similarity searches, activity prediction, etc. The database also focusses on chemical novelty and offers several ‘tranches’ for download such as ‘fragment’, ‘lead-like’, ‘drug-like’, etc.

The aim of docking HMDB, KEGG and other molecules was to determine whether there are cell metabolites or biologically relevant molecules that could fit within the region of electron density. Additionally, checks were made of the various kits and reagents from prior studies [201] in order to rule out contamination from commercial kit/molecule providers as

listed in the paper. Other molecules were docked, even those already ruled out, as negative controls (Monoolein, Hemisuccinate, Glycerol Monooleate, LKM, THPTA).

For docking with SEED, the probe library was prepared from SMILES using OpenEye’s QUACPAC [137] suite for appropriate protomer and tautomer generation. Multiple 3D conformers for each protomer were generated with OMEGA [138] and, as there are few molecules in the database with rotatable bonds, no restrictions were placed on conformer generation. AM1-BCC [139] partial atomic charges were calculated for each conformer.

Generation of HMDB and KEGG libraries for use with DOCK was completed using the internal ligand generation pipeline implemented within the DOCK 3.7 suite to generate the .db2 files required. The pipeline takes SMILES as input and uses Chem Axon [140] programs for protomer generation, CORINA 4.1 [141] for initial 3D conformer generation, AMSOL 7.1 [142] to compute atomic partial charges and ligand desolvation terms and finally OMEGA to generate multiple conformers.

For this work, the ‘lead-like’ databases were downloaded directly from the ZINC12 and ZINC15 websites.

4.2.3 Docking and pose minimisation

For SEED, both polar and apolar docking algorithms were selected and all poses to an energy of $+5.0 \text{ kcal}\cdot\text{mol}^{-1}$ were retrieved. Otherwise, program defaults regarding certain parameters (e.g. protein clashes, charge cut-offs) were used. Poses were assessed in terms of fragment and receptor desolvation energies as well as total interaction energies.

For DOCK, parameters governing van der Waals clashes were increased from defaults ($20 \text{ kcal}\cdot\text{mol}^{-1} \rightarrow 500 \text{ kcal}\cdot\text{mol}^{-1}$) i.e. poses where the vdW score exceeds this were discarded. The *matchgoal* parameter was also set higher ($1000 \rightarrow 5000$) to increase sampling of molecular orientations. Simplex minimisation was used. Sphere colouring was not used. Only the highest scoring pose was retrieved. Regarding all other parameters, program defaults were used.

For ZINC screens, the positions of all pose structures within the SMO active site were optimised using SZYBKI [156]. Optimisation parameters were set to minimise torsional strain of molecular poses using the Merck Molecular Force Field (MMFF94 [157]) with the Poisson-Boltzmann solvation model [158]. All protein residues within 3 \AA of a given pose were also minimised. Program defaults regarding the optimisation method (e.g. BFGS [159])

algorithm) were used. Estimates of various energies (e.g. overall interaction, desolvation cost) were obtained from optimised protein-ligand structures.

Chem Axon’s [140] MarvinSketch v14.12.8.0 for Linux was used for drawing and displaying chemical structures and calculating pKa curves used to evaluate docked molecules. The Cambridge Structural Database’s program, Mogul [211], was used in the evaluation of molecular torsion angles. Receiver Operator Curve (ROC) plots to evaluate docking performance were generated with Python scripts in the DOCK pipeline. Molecular Tanimoto similarity (T_c) was calculated with Open Babel [96] using the ECFP4 [94] fingerprint.

4.2.4 qPCR assay

Real-time quantitative reverse-transcription PCR reactions were performed using Power SYBR Green PCR Master Mix from Thermo Fisher/Life Technologies. Relative *Gli1* mRNA levels were calculated using the comparative ΔC_t method, reported in arbitrary units and normalized with respect to *Gapdh* mRNA levels.

Primers used were custom with respect to *Gli1* (forward primer: 5'-CCAAGCCAACTT TATGTCAGGG-3' and reverse primer: 5'-AGCCCGCTTCTTTGTTAATTTGA-3') and *Gapdh* (forward primer: 5'-AGTGGCAAAGTGGAGATT-3' and reverse primer: 5'-GTGGAG TCATACTGGAACA-3'). All experiments were repeated with two mechanical replicates and at least three biological replicates.

4.3 Results

4.3.1 Identifying the region of Fo-Fc density and mapping SMO’s allosteric pocket

Docking to the unpublished structure was completed using UCSF DOCK. Spheres were added primarily to ensure coverage of the area of electron density in the middle of the binding pocket. In all, 55 matching spheres were used.

During preparation of HMDB and KEGG datasets, some molecules ($\sim 4\%$), often those with unusual atomic combinations or unclear protonation states, failed checks within the ligand generation pipeline. Generally, these were highly-charged molecules or had a large number (< 4) of chiral centres that one would not expect to interact with SMO or be resident in any system where SMO is present so were omitted from consideration.

Upon omission of failed protomers, the finalised datasets contained 1001 protomers with respect to HMDB and 15 245 with respect to KEGG.

All poses were evaluated in terms of whether they fit within the boundaries of the CURED in terms of its complete volume. Candidate molecules that met this criteria were subsequently evaluated in terms of unsatisfied hydrogen bond donors and acceptors and shape complementarity. None of the poses retrieved by docking either HMDB nor KEGG satisfied both criteria.

The region of electron density within the TM7 bundle of the unpublished structure was divided into multiple sections (Figure 4.5). These sections reflected differences in predictions of what probes molecules were favourable and in which section they were favourable.

The entire region was divided into three major segments (Upper, Middle, Lower) and two points between them (Transition points One and Two).

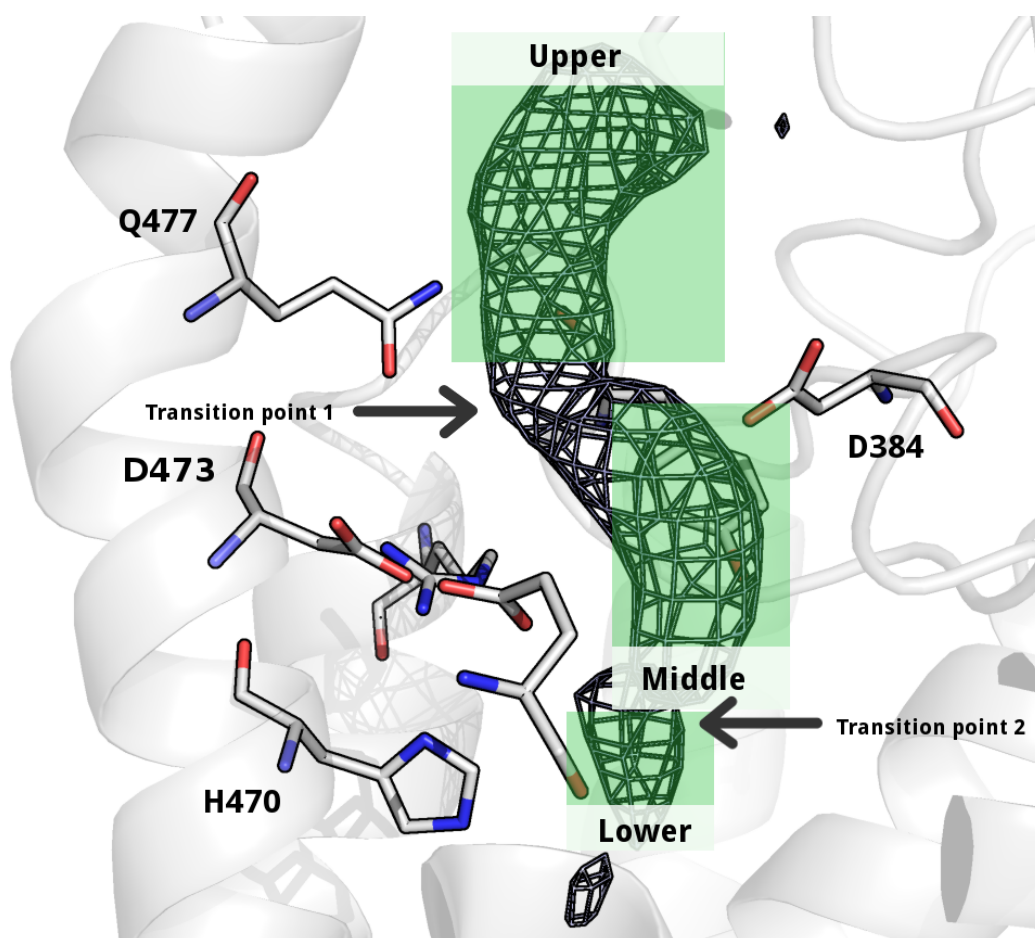


Figure 4.5: A region of Fo-Fc density (dark grey, mesh representation) from an unpublished X-ray structure is depicted, key residues (light grey, stick representation) are labelled. Indicative sections of the CURED were categorised by apparent size and shape.

Generally, highly polar molecules (e.g. methyl sulfone, propionic acid) retrieved poses with unfavourable energies. Both small polar and apolar molecules (e.g. ethanol, tert-butane) tended to result in poses with higher energies and fewer poses in general, as did non-aromatic rings. Rings found fewer favourable poses in the Upper region and those that had large receptor desolvation costs and species 6-ring sized or larger had increasing receptor desolvation costs the lower into the pocket they were placed.

As expected, highly-scored poses tended to be driven by apolar binding. Some probes (e.g. phenol, mesoinisotol) scored higher due to polar substituents interacting with polar residues. Although these substituents were generally placed outside of the CURED, their apolar interaction energies inside it were similar to non-substituted analogues (e.g. benzene).

In the Upper region, few poses addressed the upper portion of this section explicitly or fit it perfectly. However, in general, probes with large apolar sections had high-scoring poses placed here and were often unfavourable any deeper in the CURED region. This was generally driven by lower apolar interaction energies and poses placed closer to or inside the CURED tended to have lower receptor desolvation costs. Fused-ring systems tended to score higher in this section and, in some cases, were placed almost within the contour of the CURED region (Figure 4.6). The aforementioned trend of more favourable receptor desolvation also held true with poses placed closer to or within the CURED.

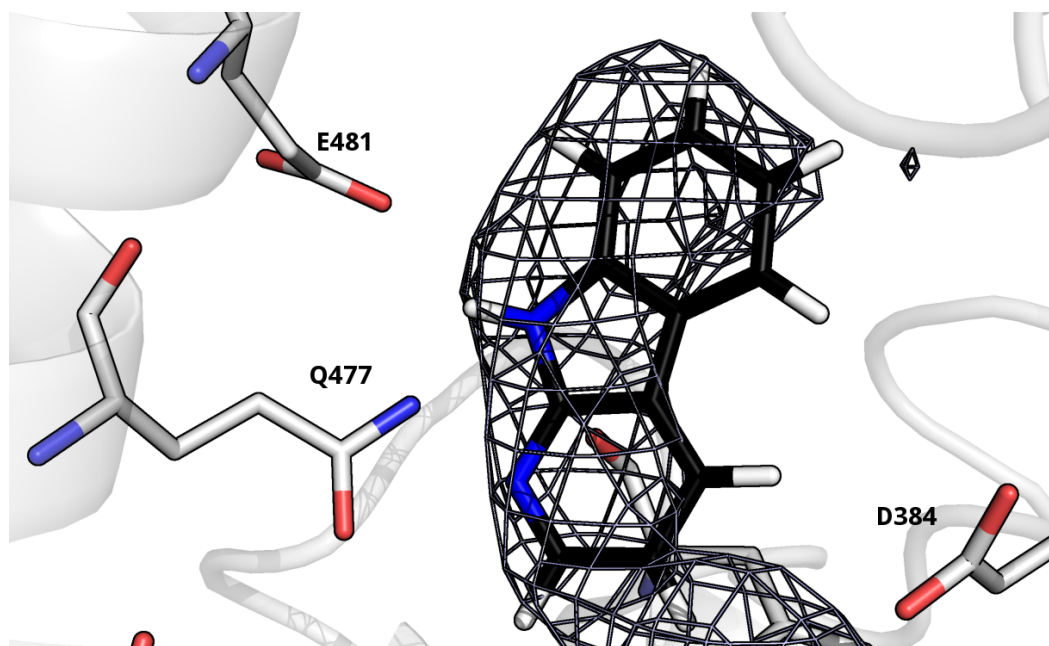


Figure 4.6: Example A-carboline docking pose (black carbons, stick representation) within the CURED (dark grey, mesh representation) in the SMO binding pocket (TM7 bundle). Key residues (light grey, stick representation) are labelled. Poses placed here tended to have lower protein desolvation and large fused-ring systems tended to score higher, indicative of greater shape complementarity with the receptor pocket.

A corner of the upper region was within hydrogen-bonding distance of E481 (Figure 4.7). Although no docking pose was directly placed here, distance measurements between this corner of the CURED and the positions of E481 and Q477 suggest a hydrogen-bond donor at this position would fit, possibly interacting with both.

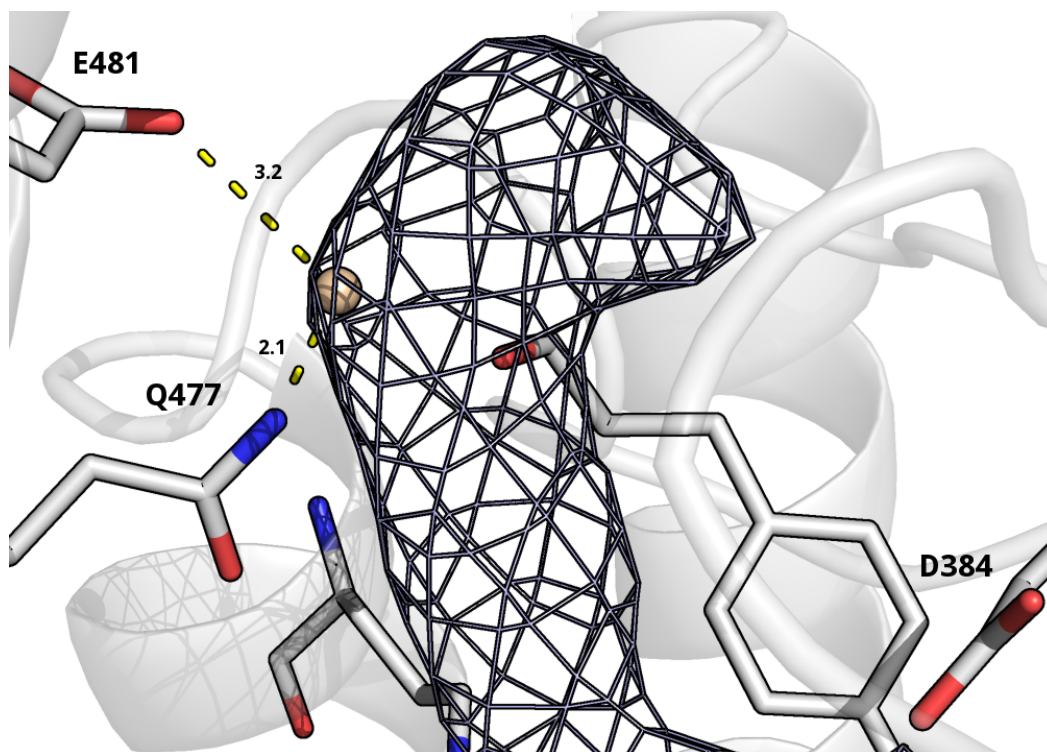


Figure 4.7: Distances (\AA) between E481 and Q477 and the CURED (dark grey, mesh representation). Key residues (light grey, stick representation) are labelled. Distances (\AA) are represented by dashed yellow lines and measured between polar atoms on residues and a sphere (straw colour) placed within the boundary of the CURED.

For the Middle region, at Transition point 1, the majority of probe poses within the CURED were unfavourable, particularly any ring-sized moieties. This is suggestive of chains linking the upper and middle sections. SEED was generally unable to place poses in the top part of this segment, particularly ring-sized probes. The main portion of this section is already very close (1.8 \AA) to D384 so favourable poses with most probe molecules, considering neighbouring residues, were unlikely (Figure 4.8).

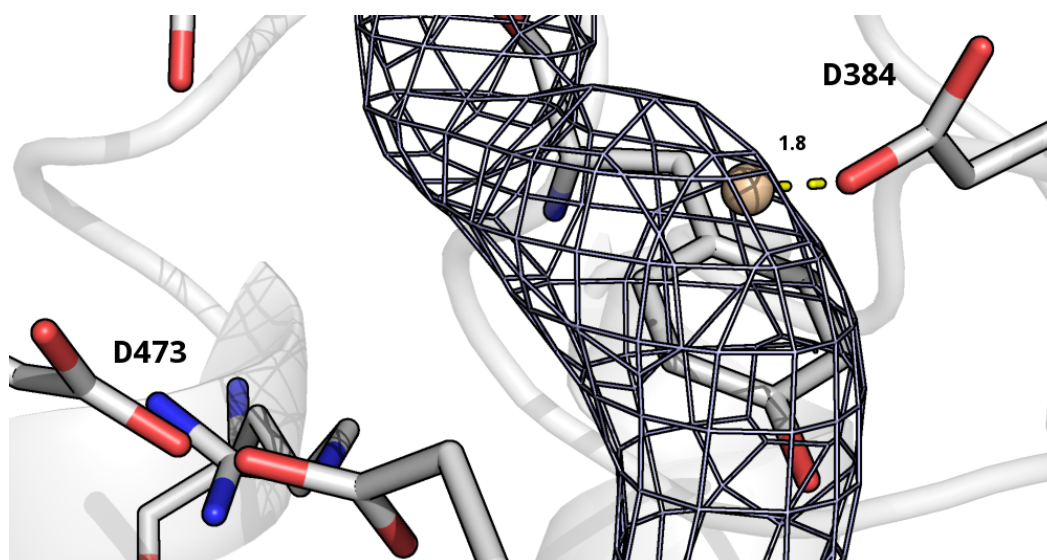


Figure 4.8: Distance (Å) between D384 and the CURED (dark grey, mesh representation). Key residues (light grey, stick representation) are labelled. Distances (Å) are measured between polar atoms on residues and a sphere (straw colour) placed within the boundary of the CURED.

An exception was methanol as several poses were found with favourable polar interaction energies between methanol and D384 (Figure 4.9). The closer poses were placed to D384, whether inside or slightly outside the CURED's boundaries, the greater the interaction energy with little increase in desolvation cost.

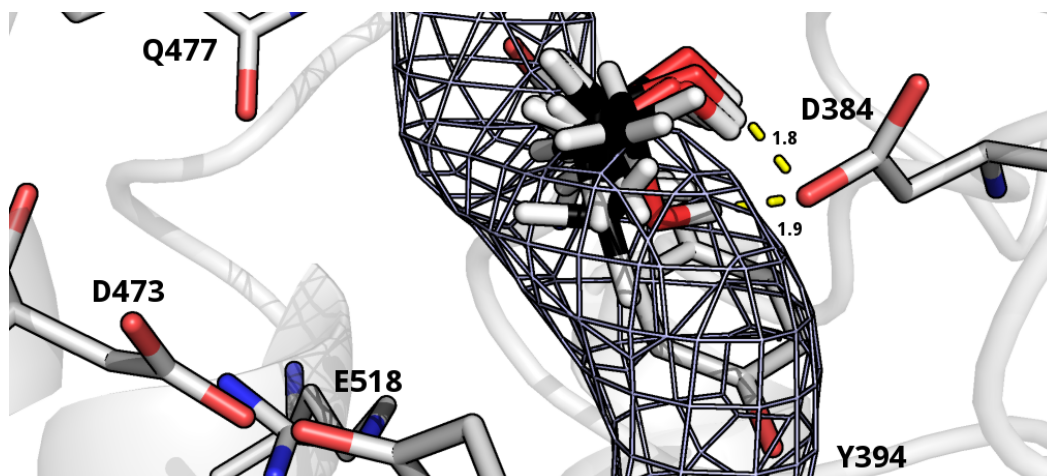


Figure 4.9: Docked methanol (black carbons, stick representation) poses within the CURED (dark grey, mesh representation) in the SMO binding pocket (TM7 bundle). Key residues (light grey, stick representation) are labelled. Predicted polar interactions are represented by yellow dashed lines and distances in Angstroms (Å). Methanol's docked positions were indicative of a hydrogen-bond donor interaction with D384 as interaction energies between poses and this residue improved the closer they were placed to D384 with little increase in desolvation cost.

Poses in the lower part of the the Middle region were retrieved with high receptor desolvation penalties and consequently unfavourable energies for any molecules that were 6-member ring-sized (e.g. benzene) or bigger (Figure 4.10). A proximal bump in the CURED was not

directly addressed by any probes but it is within range of polar residues (e.g. Y394, S287) so moieties could form polar interactions. Substituted rings (e.g. phenol) were generally unfavourable but some 5-member ring species (e.g. N-methylpyrrole) found poses that were favourable near this position.

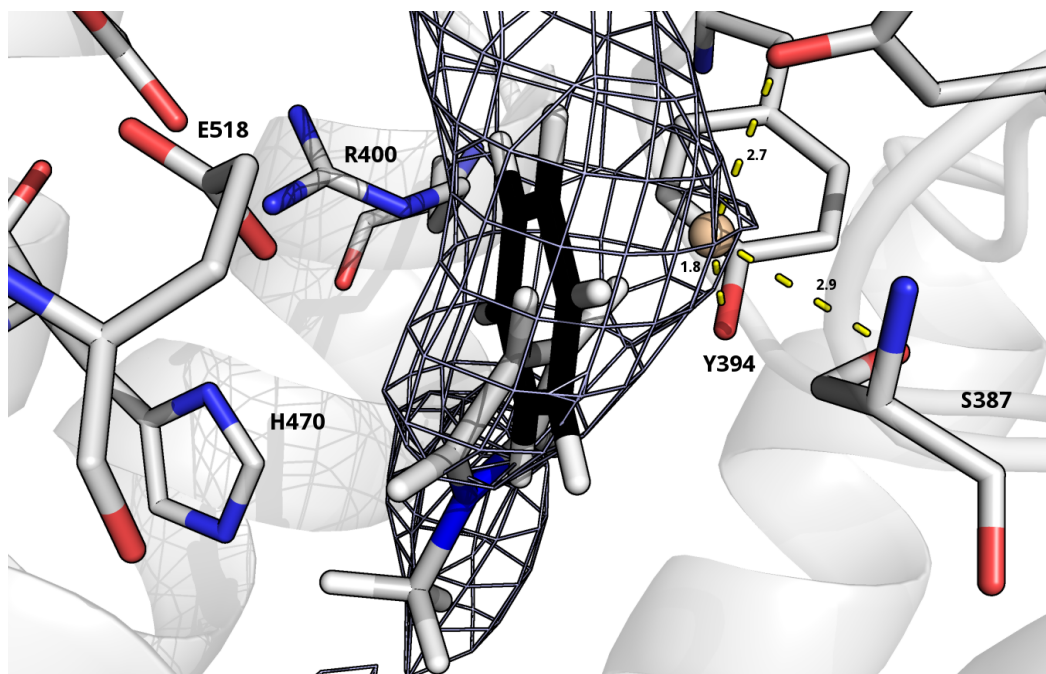


Figure 4.10: Favourable N-methylpyrrole (light grey carbons, stick representation) vs unfavourable benzene (black carbons, stick representation) in similar positions within the CURED (dark grey, mesh representation) with the distance to a proximal bump in the CURED represented by a sphere (straw colour). Distances (\AA) are represented by yellow dashed lines. This is suggestive that moieties the size of smaller endocyclic systems (or smaller) are more likely to be favourable in lower CURED regions.

Larger molecules with poses that crossed between the lower segment through Transition Point 2 and into the lower CURED region (e.g. indole) tended to be highly unfavourable and possess high receptor desolvation costs (Figure 4.11).

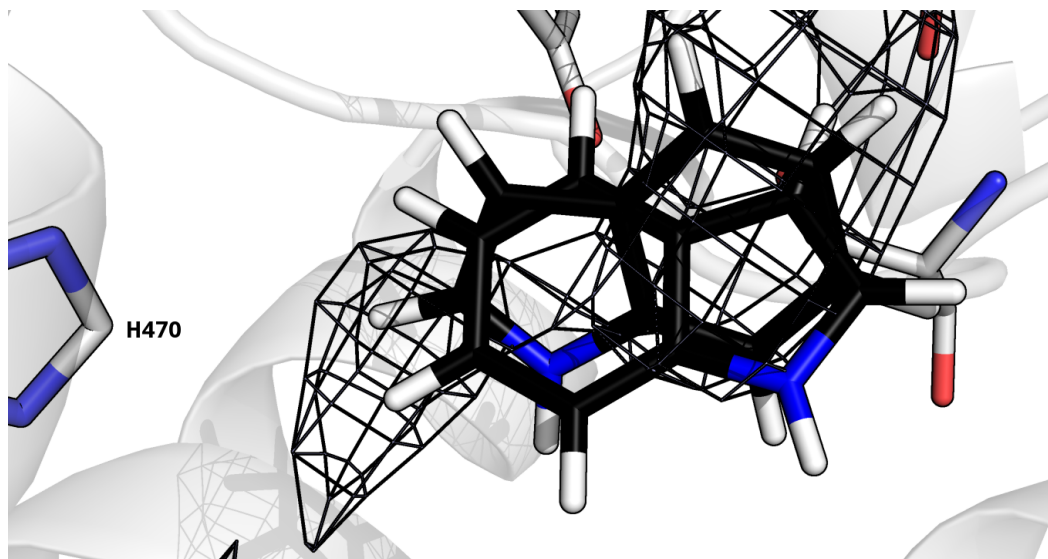


Figure 4.11: Unfavourable indole poses (black carbons, stick representation) within the CURED (dark grey, mesh representation) in the SMO binding pocket (TM7 bundle). Key residues (light grey, stick representation) are labelled. Endocyclic systems larger than benzene tended to score highly unfavourably, suggestive that moieties must fit within the CURED's density. This is as opposed to a large moiety being placed in this position where calculated atomic density is missing.

Poses from any probe that penetrated toward R400 resulted in very high receptor desolvation costs and unfavourable energies. This and the position of R400 (Figure 4.12) supports no direct interaction between residues in this area with the CURED molecule.

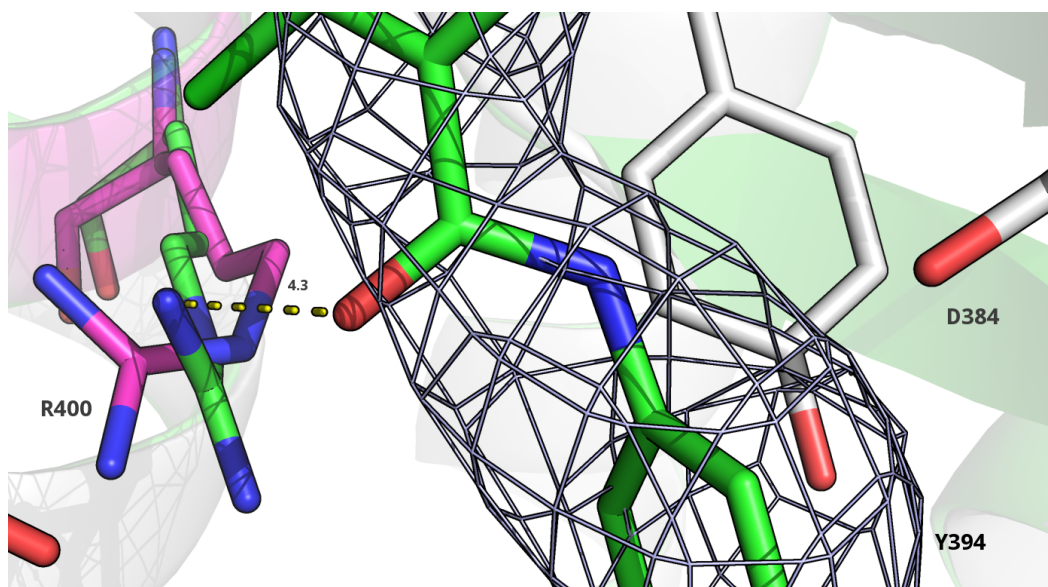


Figure 4.12: Position of R400 (magenta carbons, stick representation) in the CURED (dark grey, mesh representation) X-ray structure vs the same residue's position with respect to bound ligand Vismodegib (green carbons, stick representation). Key residues (light grey, stick representation) are labelled and distance (Å) are represented by dashed yellow lines. The distance between the crystal position of Vismodegib's amide Oxygen and R400 is indicative of no or a weak interaction. The position and orientation of R400 is further away with respect to the boundary of the CURED thus an interaction between R400 and a moiety within the CURED is unlikely.

Finally, for the Lower region, moieties placed within transition Point 2 generally had very large receptor desolvation costs and unfavourable interaction energies. Any ring-sized molecules tended to be unfavourable in this region. However, receptor desolvation costs tended to decrease for poses closer to or within the CURED.

Although a molecule could not be conclusively identified, a schema for favourable moieties from docked probes can be proposed (Figure 4.13).

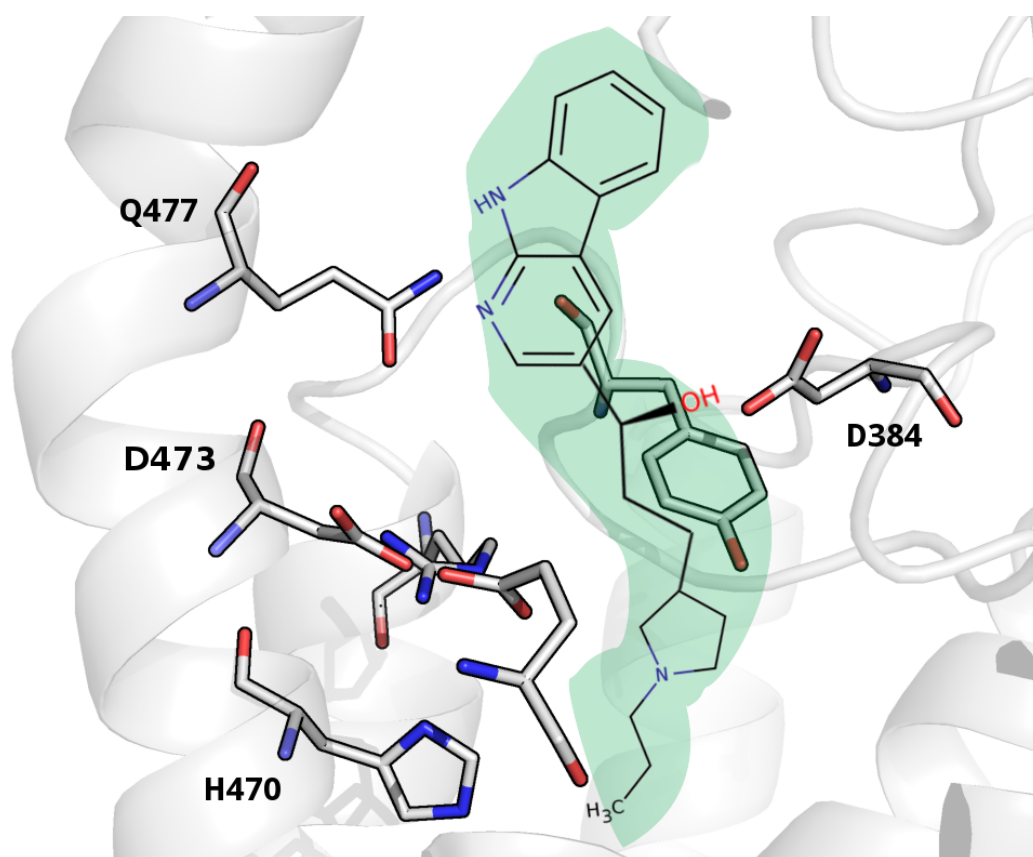


Figure 4.13: Example 2D structure of a molecule indicative of moieties favourable within the CURED (green area) in the SMO binding pocket (TM7 bundle). Key residues (light grey, stick representation) are labelled.

A complete list of all probes used and calculated interaction energies can be found in the Appendix Table 7.18

4.3.2 ZINC virtual screen - Round 1

Several constraints were applied to both the ZINC12 and ZINC15 datasets downloaded for docking. Molecules needed to be purchasable, their molecular weight between 250 and 350Da, $\log P \leq 3.5$ and contain protomers within a pH range of 6-8. Applying these constraints led to database sizes of 4 601 296 and 3 841 554 respectively.

For the structures prepared from 5L7I, spheres were added around the position of the bound

Vismodegib ligand atoms. Several were added within hydrogen bonding range of key residues such as the hydrogen bonding network between R400, H470, D473 and E518 and others that have been shown in other studies to interact with SMO ligands (N219, D384, Q477). In all, 50 matching spheres were used (Appendix Figure 7.10).

Spheres were placed near the same residues for the 4N4W X-ray structure. However, additional spheres were placed deeper in its binding pocket. This was to ensure sampling of poses placed in similar positions to the SANT1 ligand, some spheres placed as deep as L325 and M326. In all, 80 matching spheres were used (Appendix Figure 7.11).

Known actives against SMO were downloaded from the ChEMBL website. After removing duplicates, this resulted in a dataset of 676 molecules. These were used to generate 51 075 DUD-E decoys for docking against SMO.

Docking of actives and decoys with DOCK defaults initially produced worse than chance performance with the Vismodegib structure, particularly with respect to ‘early enrichment’. Early enrichment describes the situation where the correct predictions of known actives are also amongst highest ranked molecules. In general, few poses ($142/676 = 21\%$) of known actives were retrieved and increasing the size of the box around the binding site ($5 \text{ \AA}^3 \rightarrow 15 \text{ \AA}^3$) did not improve early enrichment of docked molecules.

It was determined that default DOCK 3.7 parameters regarding conformational sampling and surface clashes are set more strictly than in previous versions. The default values mean that far fewer orientations are sampled prior to energy minimisation, where poses clashing with the molecular surface are finally rejected. It was therefore considered likely that many viable poses are likely rejected at the minimisation stage. Prior work [24] has determined that increasing sampling tolerances, coupled with energy minimisation, results in monotonic improvement of ligand enrichment over decoys for most DUD-E targets.

Subsequent increases in conformational sampling and vdW score tolerances retrieved more known actives (78%) and improved both early enrichment of actives and performance in terms of logAUC (Appendix Figure 7.12, Plot A).

Docking with the ZINC12 lead-like dataset led to an initial list of 83 molecules after checks for availability, torsional strain and likely protonation states. Of these, 44 were selected after an initial visual inspection, common in high-throughput and virtual screening [166,167]. These molecules were assessed and graded by lab members experienced in docking and the grades

analysed. Subsequently, 11 were selected for purchase after checking for similarity with a prior virtual screen using the ZINC12 dataset [212]. Molecules chosen for purchase were found to be generally dissimilar to the ligands in that study (T_c : $\bar{x} = 0.22$, $\sigma_{\bar{x}} = 0.07$) and known actives (T_c : $\bar{x} = 0.32$, $\sigma_{\bar{x}} = 0.10$).

Many of the poses retrieved did reproduce the interaction between Vismodegib and D384 from its crystal structure but also tended to place apolar moieties into pockets between N219 and F484 below Y394. Polar interactions with D384 tended to be between charged amino groups rather than amide groups as per the Vismodegib X-ray structure. Some molecules (Figure 4.14) also had additional predicted interactions with residues in ECL3 (Y394) and TM6 (E518) and these were prioritised higher for testing.

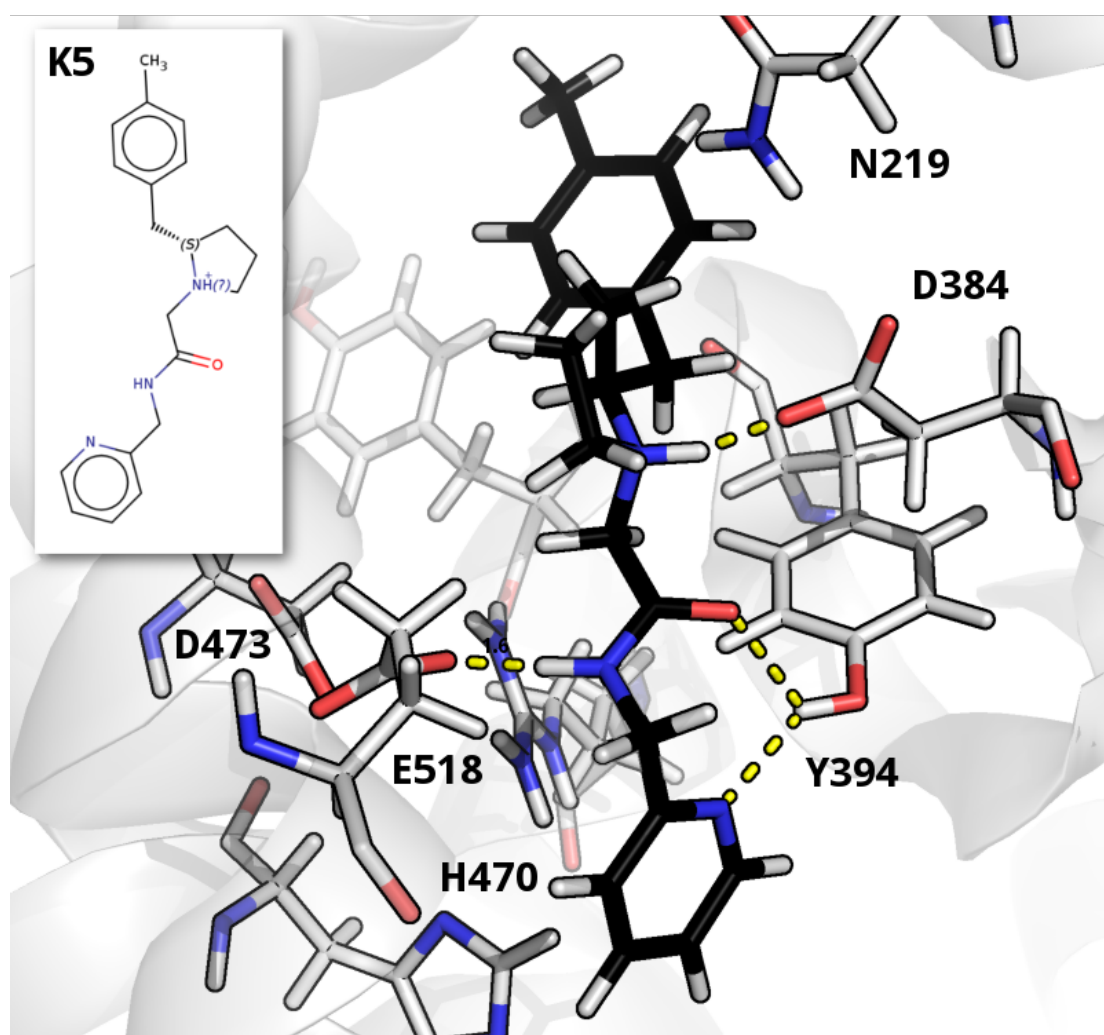


Figure 4.14: Predicted binding mode of **K5** (black carbons, stick representation) in the SMO binding pocket (TM7 bundle). Key residues (light grey, stick representation) are labelled, predicted polar interactions are represented by yellow dashed lines. Depicted here is a salt-bridge interaction between its pyrrolidine and D384 in ECL3, amide oxygen and pyridine interacting with Y394 in ECL3 and amide nitrogen predicted to interact with E518 in TM6.

The remainder of predictions came from docking to the SANT-1 structure (PDB: 4N4W).

Most compounds were predicted to recreate the binding mode displayed by SANT-1 i.e. with residues in ECL3 (Y394) and TM6 (H470) (Figure 4.15). Molecules displaying predicted interactions with residues in ECL3 (D384) and TM6 (E518) were subsequently prioritised. Amide and urea linkers were typically predicted to interact with residues in these loops and, as in common in SMO ligands, apolar moieties were placed both above and below these linkers.

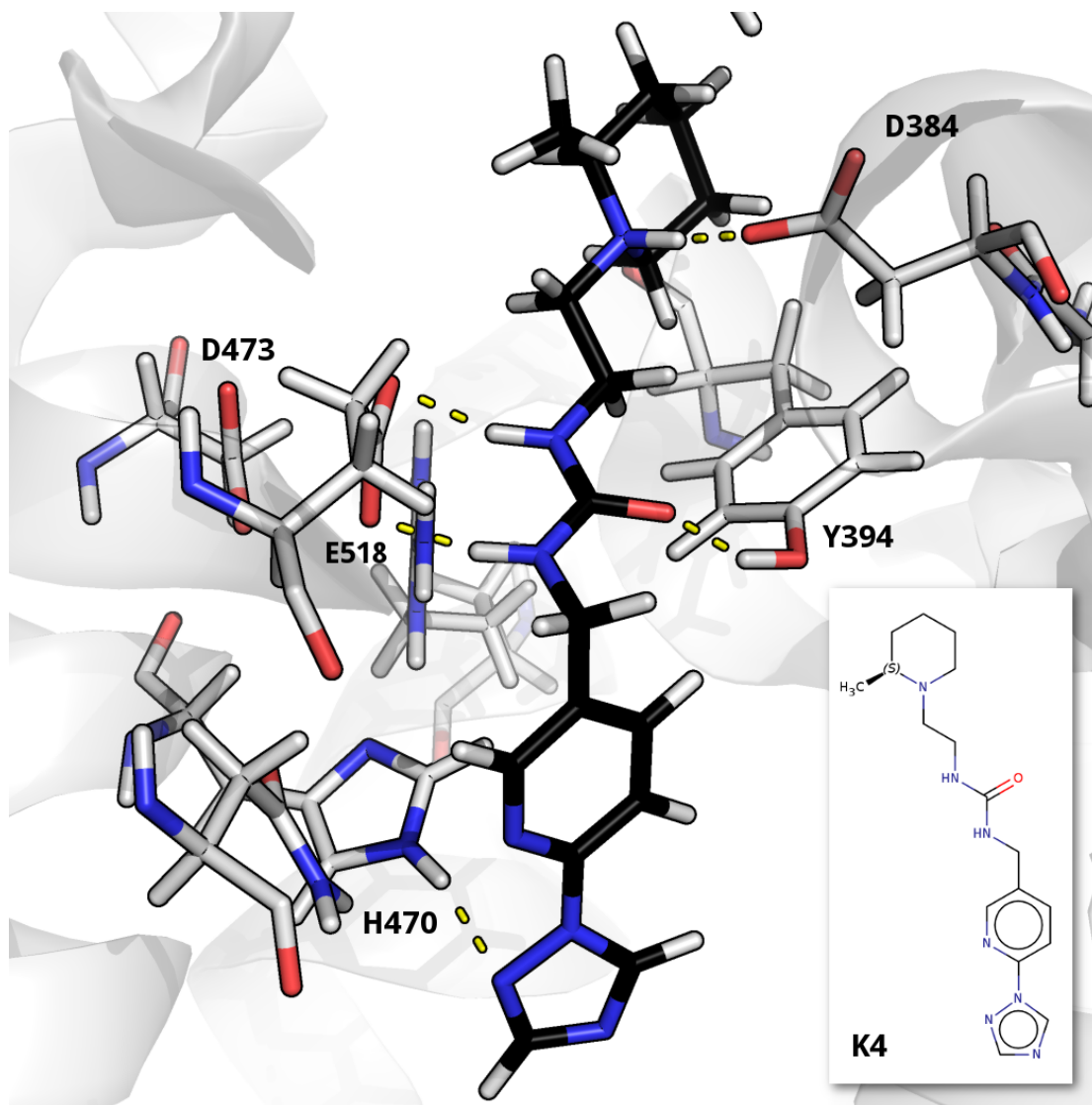


Figure 4.15: Predicted binding mode of **K4** (black carbons, stick representation) in the SMO binding pocket (TM7 bundle). Key residues (light grey, stick representation) are labelled, predicted polar interactions are represented by yellow dashed lines. Depicted here is a salt-bridge interaction between its piperidine and D384 in ECL3, urea oxygen interacting with Y394 in ECL3, urea nitrogens interacting with E518 in TM6 and triazole interacting with H470 also in TM6,

A complete list and 2D structures of molecules **K1** to **K11** can be found in the Appendix Table 4.1

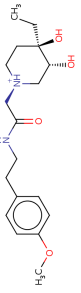
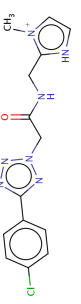
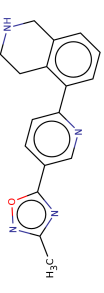
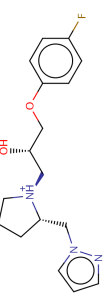
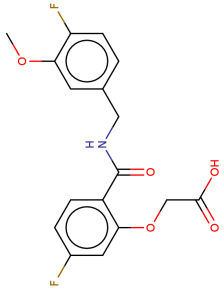
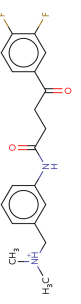
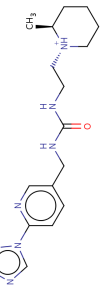
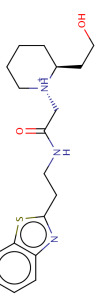
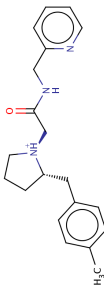
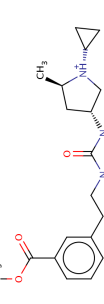
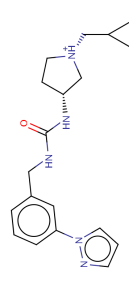
Number	ZINCID	2D Structure	Number	ZINCID	2D Structure
K1	C77522902		K7	C78985253	
K2	C71439889		K8	C71875599	
K3	ALR-75 (internal ID)		K9	C87100228	
K4	C44968732		K10	C96970747	
K5	C71904662		K11	C91704213	
K6	C95398208				

Table 4.1: Predicted Round 1 SMO inhibitors, unique identifiers from the ZINC dataset and their 2D structure. All compounds were subsequently tested for activity against SMO in cellular assays by collaborators.

4.3.3 Assay results - Round 1

Repression of Hh signalling for all 11 compounds was tested via qPCR. *Gli* mRNA is a common metric for Hh signalling activity as it is a target gene of the Hh pathway. Measurements of relative *Gli1* mRNA transcription levels were taken and normalised with respect to *Gapdh* mRNA levels. Ligands of known concentration were introduced into the assay and residual signal thus quantified a given ligand's ability to repress Hh signalling.

Assays were initially conducted at two ligand concentrations (10 μ M and 100 μ M). Measured residual transcription levels with ligand present were compared to two controls, residual transcription with no ligand present and with a potent inhibitor, SANT-1. Three compounds, **K5**, **K6** and **K9**, displayed significant concentration dependence with respect to residual transcription levels when compared to controls (Figure 4.16).

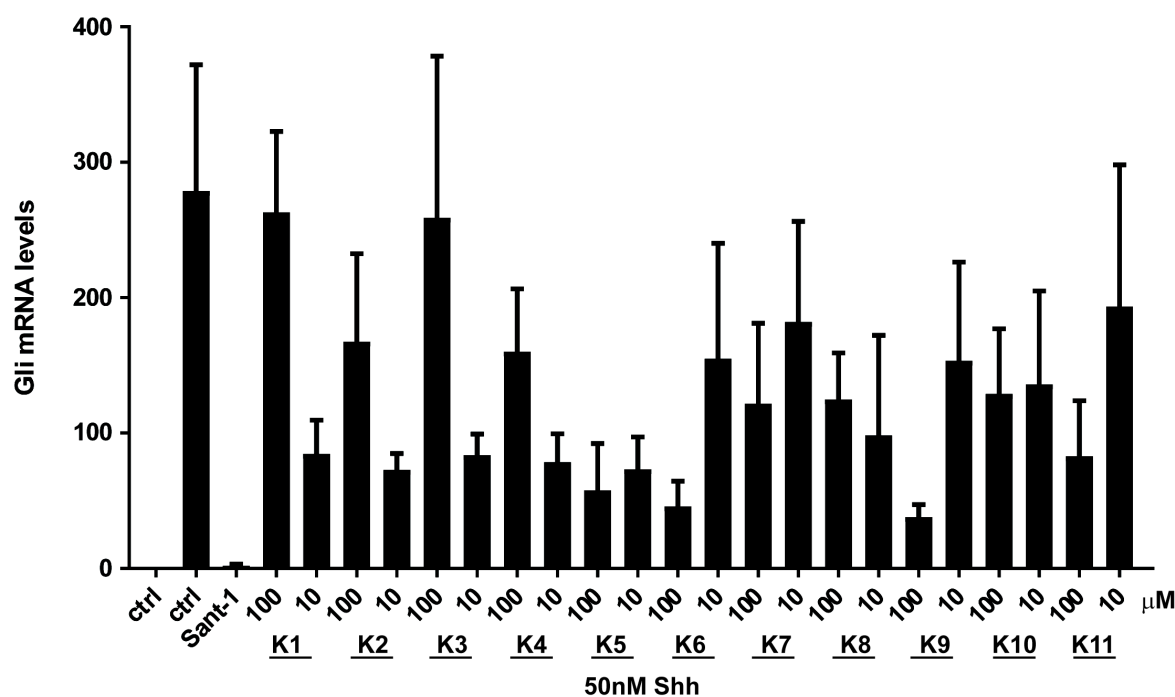


Figure 4.16: Normalised Gli mRNA levels for **K1** - **K11** at 10 μ M and 100 μ M. All compounds were tested with positive (SANT-1) and negative (no ligand) controls and reflect the constitutive activity of SMO. Statistical analysis of Gli1 mRNA levels across samples was performed via one-way ANOVA with Holm-Sidak post-hoc tests for multiple comparisons. Ligands **K5**, **K6** and **K9** displayed significant ($p < 0.01$) concentration dependence with respect to residual mRNA transcription levels.

4.3.4 Minimisation of the binding pocket found more SMO poses

Relatively few poses were found in the first round of docking to the Vismodegib X-ray structure (PDB: 5L7I) than would be expected for a drug-bound structure. It was apparent from visual inspection of poses retrieved from docking that the dominant chemotype retrieved

was pyrimidines with unlikely protonation states ($\sim 26\%$ of the top-2000 poses). These poses were generally placed near the hydrogen-bonding network (R400/H470/D473/E518) and no lower.

There was also some degree of excessive intramolecular strain in poses that was unable to be resolved by ligand minimisation, even when minimising nearby ($\leq 3 \text{ \AA}$) residues as well. Primary amides in particular displayed significant torsional strain, pushing the carbonyl oxygen and amine hydrogen out-of-plane (Figure 4.17). Both phenomena were likely caused by large van der Waals forces imposed upon the ligand by apolar residues in the binding pocket and gave rise to the notion that perhaps residues in the binding pocket itself were not fully relaxed.

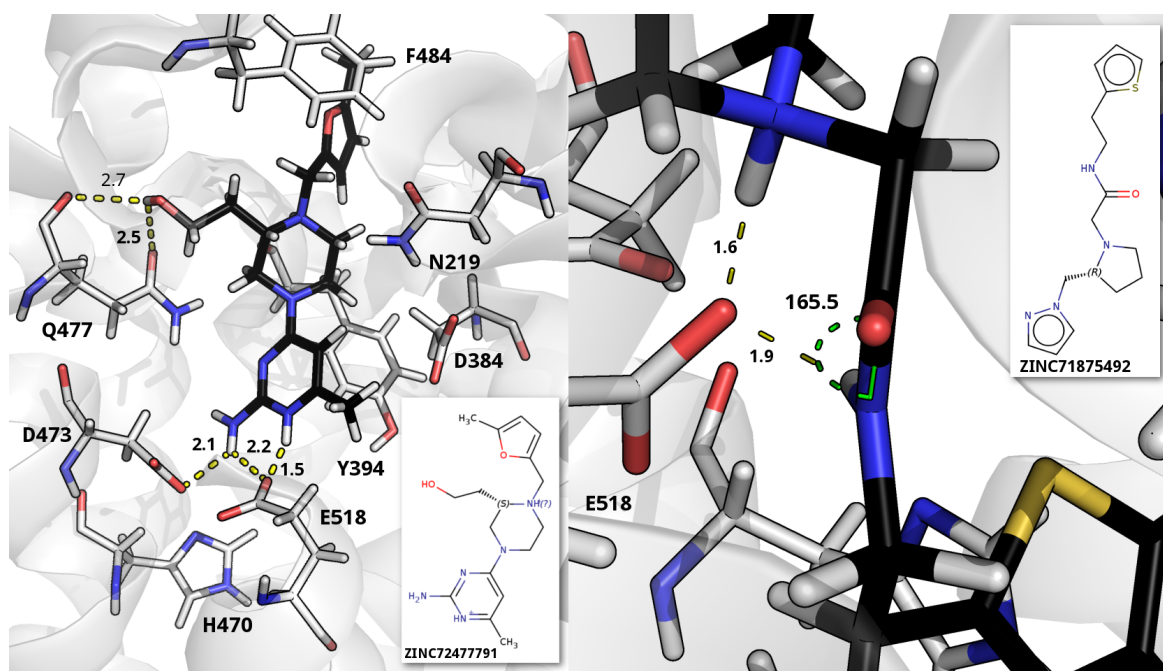


Figure 4.17: Example molecular poses from docking with unlikely pyrimidine protonation state (left, black carbons, stick representation) and primary amide with torsional strain (right, black carbons, stick representation) in the SMO binding pocket (TM7 bundle). Key residues (light grey, stick representation) are labelled, predicted polar interactions are represented by yellow dashed lines, distances (\AA) and angles were measured. It was considered that the placement of poses in this manner, even after energy minimisation of both the ligand and surrounding residues was conducted, was suggestive of large vdW forces being applied to the molecules being docked. Additionally, the degree of deviance from 180° of some amides, although within the boundaries of previously published X-ray structures, was considered excessive. This was considered *prima facie* evidence that the pocket should be relaxed via energy minimisation to a lower energy conformation in order to find more likely docking poses.

To find more poses and minimise intramolecular strain, all residues within 8 \AA of Vismodegib's position in the binding pocket were minimised - 66 in total (Figure 4.18). Most residues within this distance were in TM6 and the rest spread evenly between Extra-Cellular

Loops 1 and 3. Residues were minimised using CHARMM and MOE and the differences between these results compared to the X-ray structure and to each other.

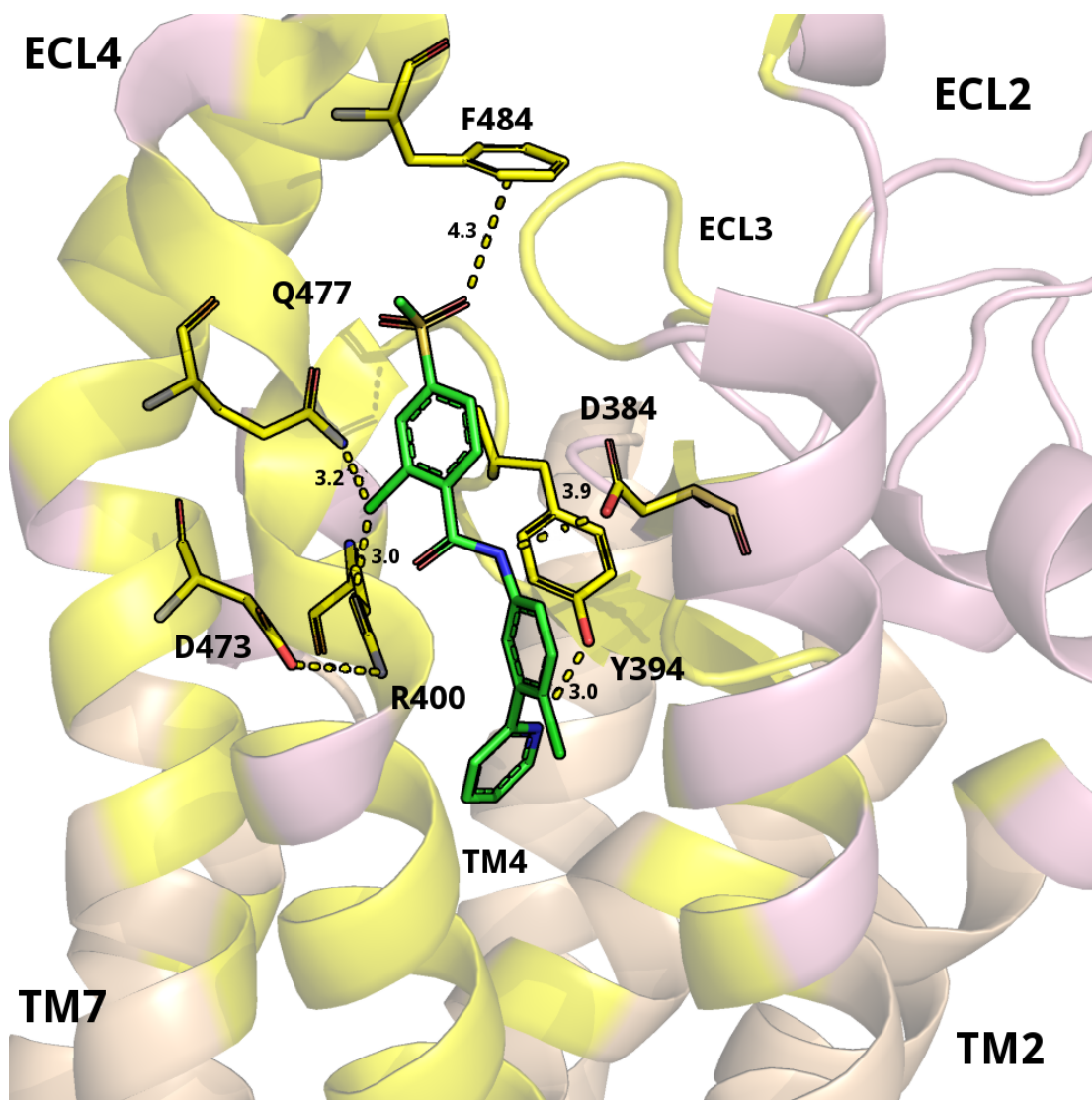


Figure 4.18: Depiction of residues (banana yellow, cartoon representation) within 8 Å of the binding position of Vismodegib (green carbons, stick representation) in ECL1, 3 and 4 and TM6 (pink, cartoon representation) depicting those that were relaxed by minimisation with both CHARMM and MOE. Key residues (yellow, stick representation) are labelled, distances (Å) are represented by yellow dashed lines.

A complete list of minimised residues can be found in Appendix Table 7.19

Overall, there was little movement of binding pocket residues. Movement of atomic positions in terms of RMSD for each residue was compared to the original X-ray structure. There were lower RMSDs from the X-ray structure for the CHARMM-minimised structure (RMSD: $\bar{x} = 0.40$, $\sigma_{\bar{x}} = 0.21$) when compared to the MOE-minimised structure (RMSD from X-ray structure: $\bar{x} = 0.54$, $\sigma_{\bar{x}} = 0.36$). Movements in the latter were, however, more ‘inward’ resulting in a more constricted binding pocket (Figure 4.19). There was, accordingly, a lower overall

energy for the structure minimised by CHARMM ($-18\,341.69\text{ kcal}\cdot\text{mol}^{-1}$) when compared to that obtained from MOE ($-18\,197.09\text{ kcal}\cdot\text{mol}^{-1}$). Consequently, the CHARMM-minimised structure was chosen for further docking.

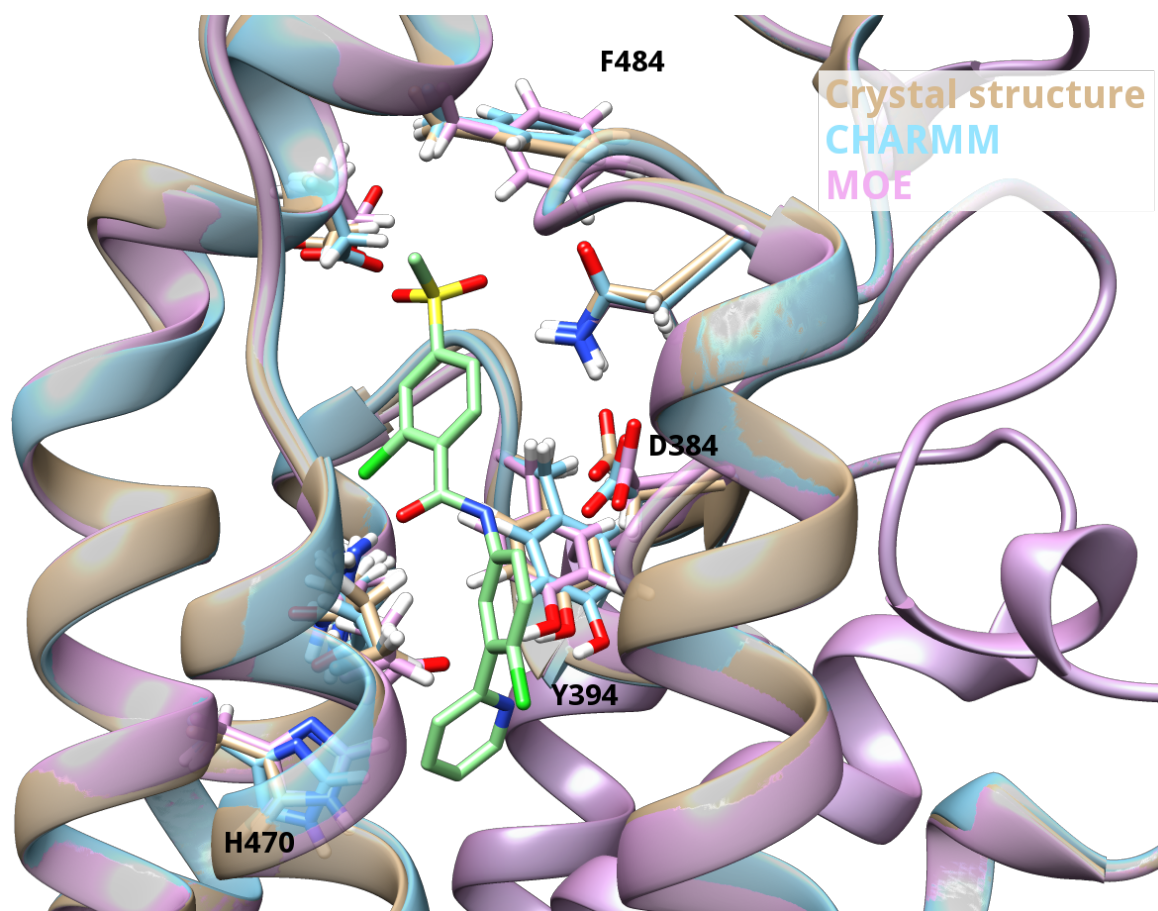


Figure 4.19: Selected minimised binding pocket residues comparing the crystal structure and after minimisation with CHARMM (light blue, cartoon representation) and MOE (pink, cartoon representation) with the binding position of Vismodegib (green, stick representation) for reference. Key residues (stick representation) are labelled. Minimisation with CHARMM resulted in a lower energy structure when compared to MOE and more outward movement of residues. The resultant structure was a more ‘open’ binding pocket and was used for further docking.

Descriptive statistics for the RMSDs and complete CHARMM output of the energy calculation can be found in Appendix Tables 7.20 and 7.21 respectively

As in Round 1, known actives against SMO and decoys were redocked to the relaxed structure to assess its performance. More poses (98% of known actives) were retrieved than the unrelaxed structures and there were mild improvements both in AUC and early enrichment of actives (Appendix Figure 7.12, Plot B).

The ZINC12 dataset was also redocked to the CHARMM-minimised structure and the top-2000 scored poses were retained for analysis. Docking to the CHARMM-minimised structure

resulted in a very different top-2000 when compared to that from docking to the X-ray structure. Only 440 (22%) molecules were common to the top-2000 of both docks. Of those, there were relatively small changes in score but the same molecules displayed very large changes in ranking (Table 4.2).

	min	Q1	\bar{x}	Q3	max	Pearson r/MCC
rank	-1814.0	-669.5	-141.0	315.0	1895.0	0.08/0.08
score	-21.420	-7.305	-5.645	-3.790	3.910	0.56

Table 4.2: Five-number summary statistics (min, Q1, median, Q3, max) of differences in ranking and score for the 440 molecules common between ZINC12 dockings. Pearson correlation and Matthews Correlation Coefficient (MCC) also shown. The data confirm large differences in docking rank and extremely weak correlation (0.08) when comparing molecules docked to unrelaxed and relaxed structures. A similar but more mild effect was observed for docking score (MCC=0.56).

Additionally, the proportion of pyrimidines with unlikely protonation states decreased ($\sim 26\% \rightarrow \sim 7\%$ of the top-2000 poses) and more poses were retrieved interacting with residues in ECL1 (N219) found in previous ligands against SMO.

Though more poses were retrieved in Round 1 docking to the SANT-1 structure, it was apparent that none of them were ‘deep’ binding ligands. Tartering of residues was therefore applied to the already-prepared deep binding X-ray structure (Figure 4.20).

In all, 9 residues were tarted, primarily by increasing the polarity of the backbone carbonyl groups in apolar residues. A complete list of tarted residues and affected atoms and can be found in Appendix Table 7.22.

As with the Vismodegib structure, increasing pose sampling improved early enrichment. However, docking actives and decoys to the tarted structure resulted in worse than chance prediction for high ranked compounds and a late enrichment of known actives to SMO (Appendix Figure 7.12, Plot D). This is likely due to a large increase in electrostatic potential caused by tarting >5 residues in the same binding pocket, even apolar residues such as T466 and F462. Primarily, this meant that more polar decoys that scored poorly when docked to the untarted receptor were scored far more favourably in the tarted receptor.

Six of the known actives to SMO, also more polar molecules, scored highly when docked to the tarted receptor after being scored as unfavourable against the untarted receptor. When omitted from analysis, scores for the remaining actives that retrieved docked positions (463)

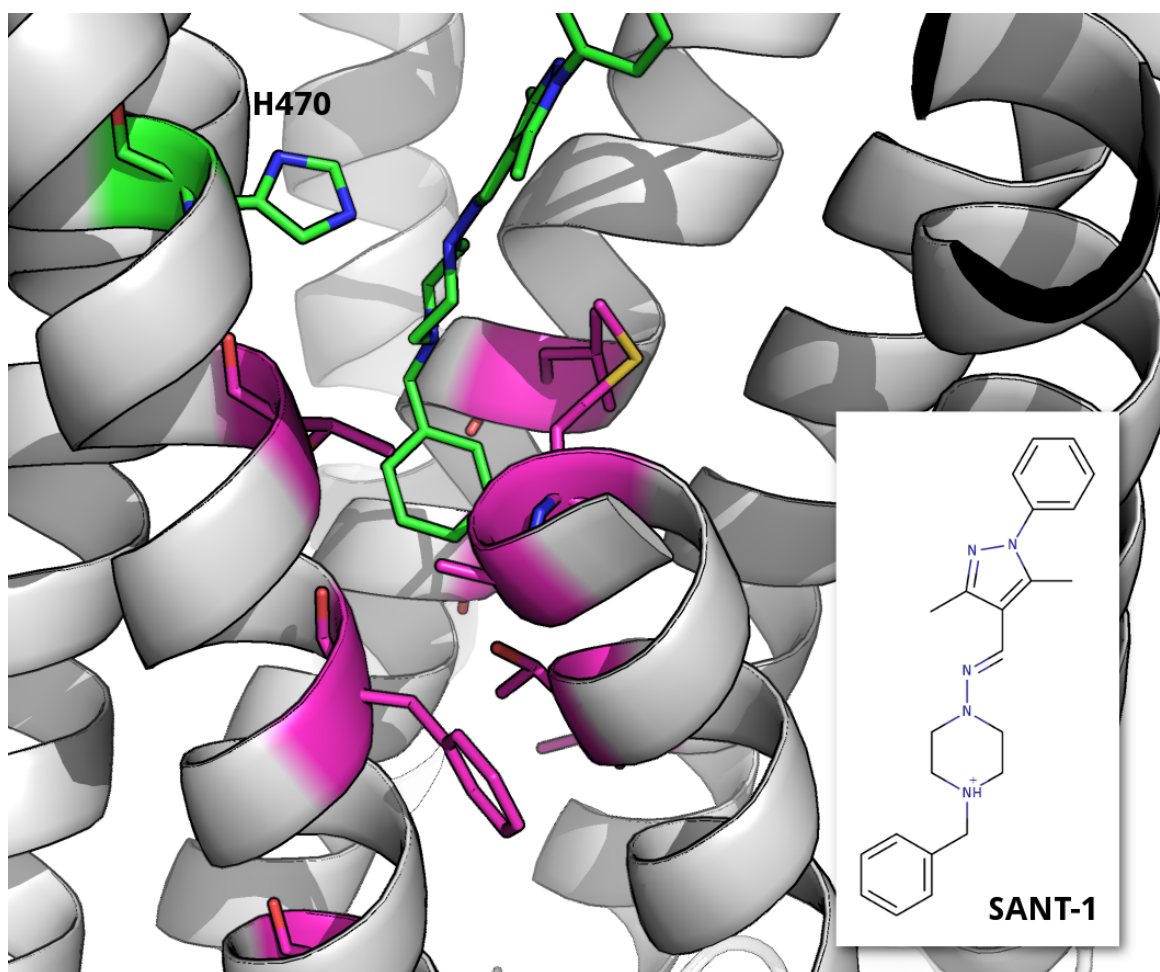


Figure 4.20: Residues (magenta, stick representation), generally apolar in nature, that were ‘tarted’ below H470 (green carbons, stick representation) in the SMO binding pocket (dark green, cartoon presentation) with the position of ligand SANT-1 (green carbons, stick representation) for reference with 2D structure (inset) displayed. In all, 9 residues were ‘tarted’ prior to redocking the ZINC12 dataset but the effect on docking score and rank was mild.

in tarted and untarted receptors were broadly correlated (Pearson $r = 0.76$). 2D structures and scores for these molecules can be found in the Appendix (Table 7.23)

Upon redocking the ZINC12 dataset, most poses retrieved (1699/85%) were common to the top-2000 poses of both tarted and untarted receptors and both ranking and scores did not change significantly (Table 4.3).

Energy minimisation of molecules docked to the Vismodegib structure revealed that using the relaxed receptor pocket did result in molecules with more favourable protein-ligand vdW energies as well as slightly more favourable protein and ligand desolvation energies. However, the mean estimated interaction energies of both rounds of molecules were very similar for those docked to the X-ray structure ($-17.45 \text{ kcal}\cdot\text{mol}^{-1}$) when compared to the relaxed structure ($-17.35 \text{ kcal}\cdot\text{mol}^{-1}$).

	min	Q1	\tilde{x}	Q3	max	Pearson r/MCC
rank	-1506.0	-96.0	-3.0	103.5	1366.0	0.87/0.79
score	-7.55	-0.17	0.14	0.46	7.95	0.94

Table 4.3: Five-number summary statistics (min, Q1, median, Q3, max) of differences in ranking and score for the 1699 molecules common between ZINC12 dockings. Pearson correlation and Matthews Correlation Coefficient (MCC) also shown. Although some docked poses displayed dramatic differences in ranking and score, generally the effect of tarting residues on docking results was mild as there were strong associations between docking to untarted and tarted receptors for rank (MCC=0.79) and score (MCC=0.94) respectively.

The effect of the tarted receptor was less pronounced. As expected, Coulomb energies were more favourable but mean overall protein-ligand interaction energies were similar for molecules docked to the X-ray structure ($-12.15 \text{ kcal}\cdot\text{mol}^{-1}$) as compared to the tarted ($-12.51 \text{ kcal}\cdot\text{mol}^{-1}$) receptor.

Generally, estimated interaction energies were higher for molecules docked to the 4N4W structure when compared to the Vismodegib structure. This is likely to be due to the higher ($\sim 10 \text{ kcal}\cdot\text{mol}^{-1}$) cost of protein desolvation for poses placed deeper in the 7TM bundle.

A summary of energies as estimated during minimisation can be found in Appendix Table 7.26

4.3.5 ZINC virtual screen - round 2

Docking of the ZINC12 and ZINC15 lead-like datasets as well as a Pharmacognosy library led to an initial list of 115 molecules after checks for availability, torsional strain and likely protonation states. Of these, 70 were again subjected to visual inspection and analysis and submitted for evaluation and grading by lab members prior to the collation of a final list. Finally, 16 were selected for purchase after again checking for similarity with La Croix et al and molecules from Round 1. Molecules chosen for purchase were found to be generally dissimilar to the ligands in that study, broadly dissimilar to Round 1 molecules and all known actives against SMO (Appendix Table 7.24).

When compared to compounds from Round 1, docking to the Vismodegib relaxed structure resulted in molecules with larger, ring-sized moieties addressing residues near the hydrogen-bond network and linking ECL3 (Y394, R400), TM6 (E518, H470) and ECL1 (N219). Again, residues in ECL3 and TM6 prone to mutations in BCCs (D473, E481) were avoided. Molecules from the ZINC15 dataset, broadly, were more likely to be comprised of

both aromatic and non-aromatic endocyclic amines so whilst several predictions included the amide/urea linker as seen in Round 1 poses, several poses were retrieved with the added benefit of rigidity and shape complementarity (Figure 4.21).

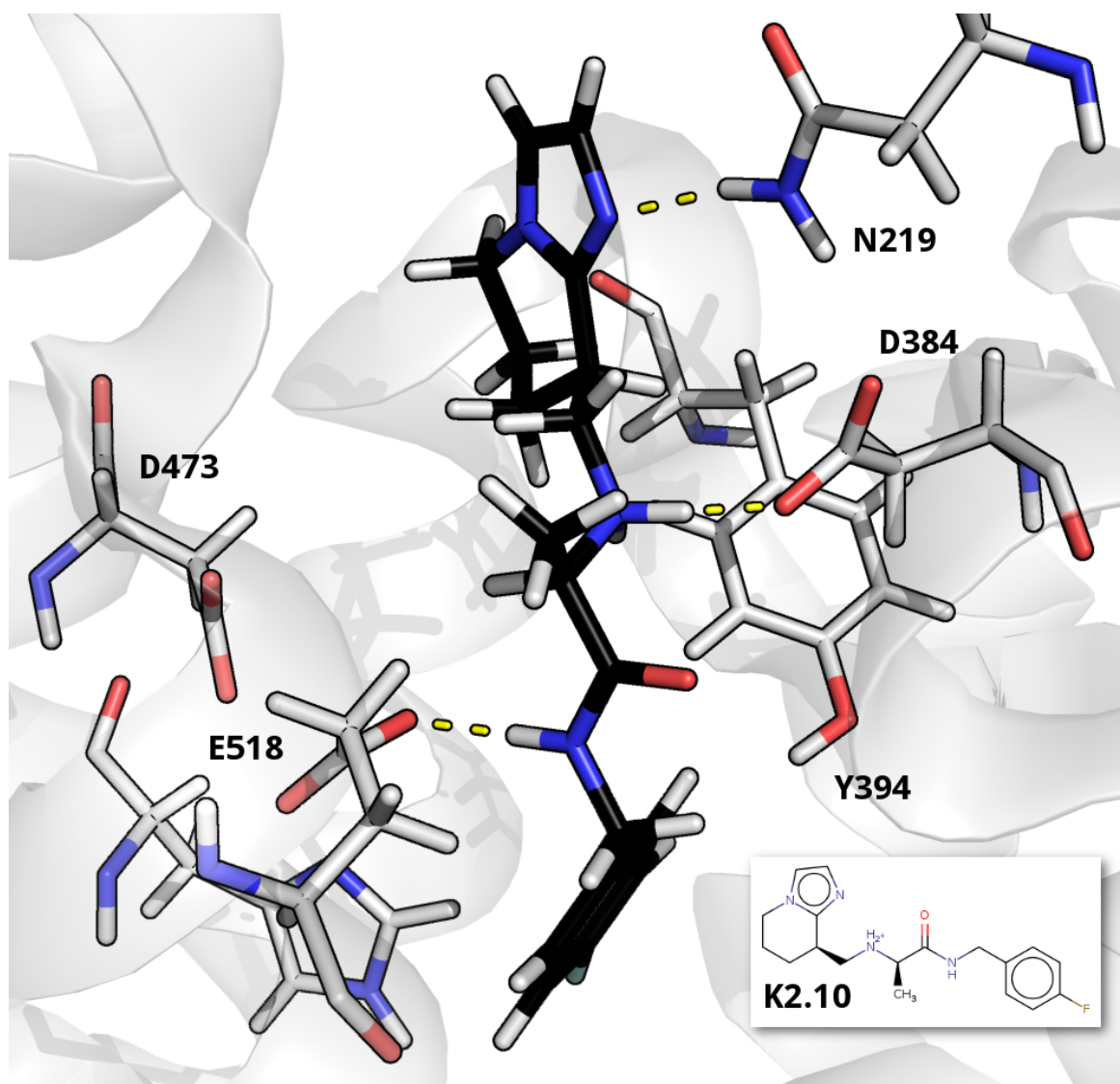


Figure 4.21: Predicted binding mode of compound **K2.10** (black carbons, stick representation) in the SMO binding pocket (TM7 bundle). Key residues (light grey, stick representation) are labelled, predicted polar interactions are represented by yellow dashed lines. Depicted are interactions between the imidazole of **K2.10** and N219 in ECL1, a salt-bridge between a primary amine and D384 in ECL3 and an interaction between its amide nitrogen and E518 in TM6.

Predictions from the tarted SANT-1 structure recreated the two polar interactions from the SANT-1 X-ray structure in ECL3 (Y394) and TM6 (H470). Additional interactions were predicted with other residues in TM6 (E518 and N521). More poses were retrieved with improved shape complementarity and more poses penetrated deep into the binding pocket when compared to poses from Round 1 (Figure 4.22).

A complete list and 2D structures of molecules **K2.1** to **K2.20** can be found in the Table 4.4

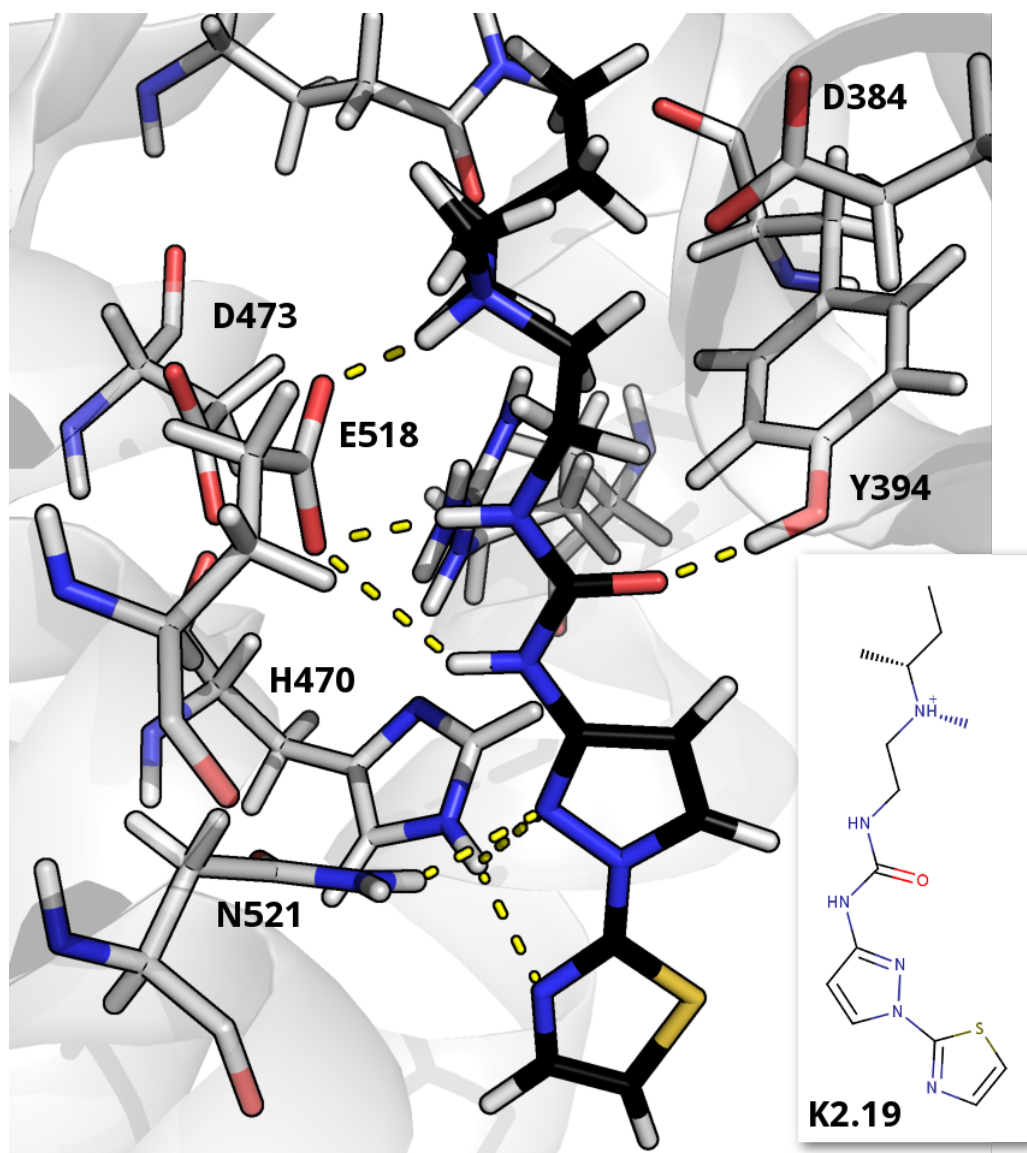


Figure 4.22: Predicted binding mode of compound **K2.19** (black carbons, stick representation) in the SMO binding pocket (TM7 bundle). Key residues (light grey, stick representation) are labelled, predicted polar interactions are represented by yellow dashed lines. Depicted is a salt-bridge interaction between the charged primary amine as well as urea nitrogens of **K2.19** with E518 in TM6, an interaction between its urea carbonyl and Y394 as well as interactions between its pyrazole and N521 and its thiazole and H470 in TM6. This was considered an ideal pose as interactions bridging residues in ECL3 (Y394) and TM6 (E518) were predicted in addition to an interaction with H470 and overall shape complementarity

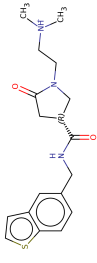
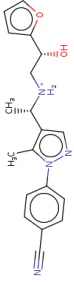
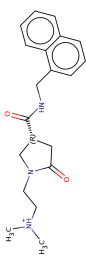
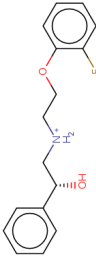
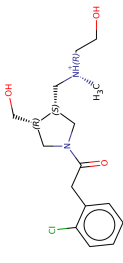
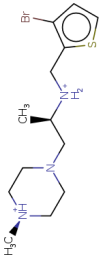
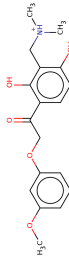
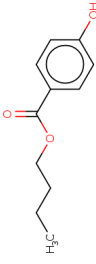
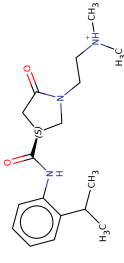
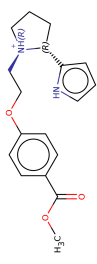

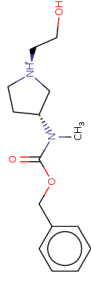
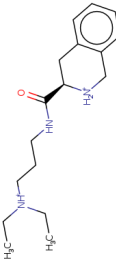
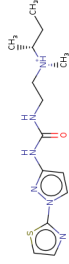
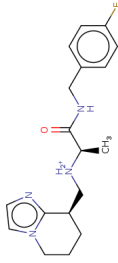
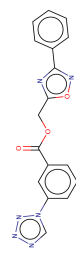
Number	ZINCID	2D Structure	Number	ZINCID	2D Structure
K2.1	C95364667		K2.11	ZINC00019553449	
K2.2	C72881007		K2.12	ZINC000002643649	
K2.3	C91877256		K2.13	ZINC000097378275	
K2.4	C11853593		K2.14	94-26-8 (CAS. No.)	
K2.5	C74195050		K2.17	C40154532	
K2.7	ZINC000069969674		K2.18	C79410921	
K2.9	ZINC000019470498		K2.19	C95444161	
K2.10	ZINC000084001097		K2.20	ZINC000008395359	

Table 4.4: Predicted Round 2 SMO inhibitors, unique identifiers from the ZINC dataset and their 2D structure. All compounds were subsequently tested for activity against SMO in cellular assays by collaborators.

4.3.6 Assay results - Round 2

Repression of Hh signalling for all 16 compounds was again tested via qPCR (Figure 4.23). Assays were again conducted at two ligand concentrations ($10\mu\text{M}$ and $100\mu\text{M}$). Of these, compounds **K2.2**, **K2.4**, **K2.7**, **K2.10-K2.14** and **K2.18-K2.20** displayed significant concentration dependence with respect to residual mRNA transcription levels when compared to controls (Figure 4.23)

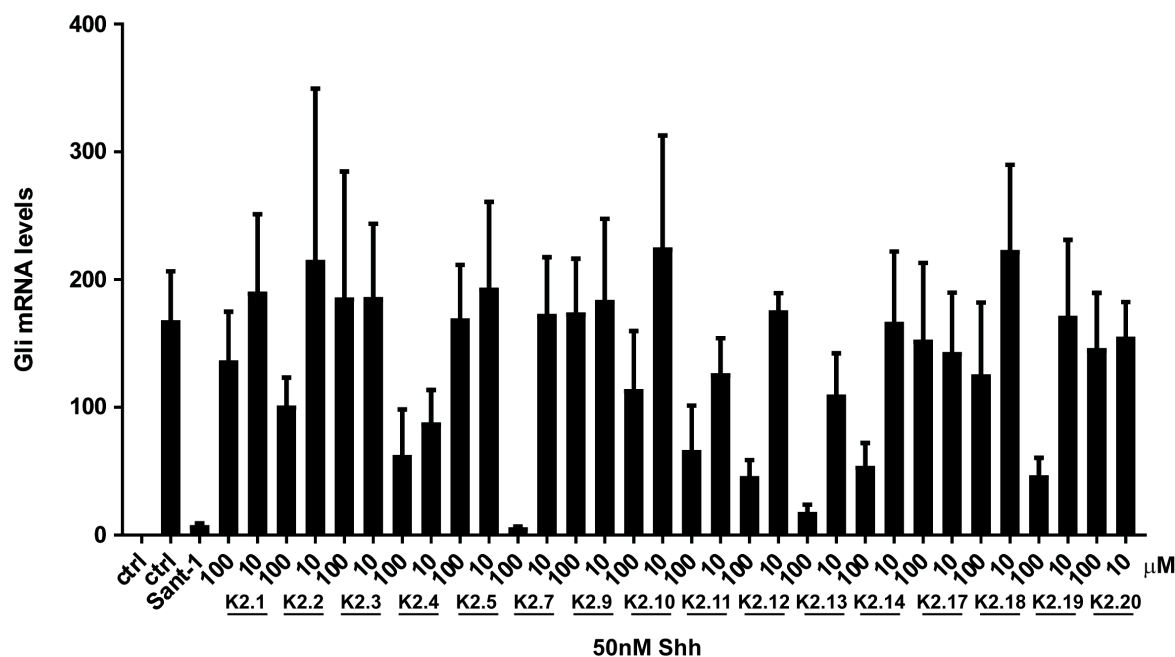


Figure 4.23: Normalised Gli mRNA levels for **K2.1 - K2.20** at $10\mu\text{M}$ and $100\mu\text{M}$. All compounds were tested with positive (SANT-1) and negative (no ligand) controls and reflect the constitutive activity of SMO. Statistical analysis of Gli1 mRNA levels across samples was performed via one-way ANOVA with Holm-Sidak post-hoc tests for multiple comparisons. Ligands **K2.2**, **K2.4**, **K2.7**, **K2.10-K2.14** and **K2.18-K2.20** displayed significant ($p < 0.01$) concentration dependence with respect to residual mRNA transcription levels.

4.4 Discussion and conclusions

Although molecular diversity was the goal with this work, it is prudent to consider which parts of a given compound should be prioritised for novelty. Common to ligands in this work was a short chain with small functional groups (e.g. aliphatic amide, urea) interacting via a hydrogen-bonding network in the middle of the binding pocket, linking and potentially stabilising its conformation. Above and below this central stability were a diverse array of aromatic groups with no clear molecular trend in terms of size, chemotype or binding mode.

Though the molecule in the CURED structure was unable to be formally identified, the

role and binding positions of endogenous regulators of SMO activity may be related and similar. Initial probe docking provided a framework for which moieties were unfavourable and where in the binding pocket and also provided useful data to inform molecule selection for subsequent screening efforts. However, docked positions of probes where interaction energies were favourable allowed the suggestion of a representative 2D structure. Essentially, a bulky endocyclic system binding near the opening to the binding pocket with possibly an aliphatic or at least less bulky tail penetrating deeper into the binding pocket. The crystallographic positions of D384, Q477 and E481 support at least one polar interaction with the ligand.

Molecules in the class of oxysterols could therefore be candidate molecules for further analysis. At present, the only endogenous regulator of SMO activity known is cholesterol [213]. And, as reported, it was crystallised in the CRD [201], not in the heptahelical bundle where the CURED is positioned. However, it remains an unresolved question whether other proteins that interact indirectly with SMO (e.g. PTC1) via cholesterol or cholesterol-like molecules regulate access to the CRD, the TM7 bundle or, indeed, both of these regions of SMO [214]. Although cholesterol itself was ruled out as a possibility for the CURED, long, cyclic endogenous ligands to SMO were known some time ago [215] and ligands with bulky fused-ring systems, such as Cyclopamine, have been crystallised with SMO in the TM7 [216] (Figure 4.24, A and B).

More recently, a class of cilia-associated oxysterols were reported to directly activate SMO and promote accumulation in the primary cilia, allowing SMO to participate in downstream signalling [217] (Figure 4.24, C-F). Although this study isolated oxysterols in cilia from non-human SMO, the authors were able to demonstrate that oxysterols were not only able to activate SMO but that they were able to activate SMO via binding in the TMD bundle and that enzymes involved in cilia-associated oxysterol biosynthesis would be enriched in domains of active Hh signaling. Which oxysterols or even other lipidic inhibitors SMO encounters before or after entering the primary cilia remain open questions. However, the broad molecular shape of oxysterols when compared to the shape of the CURED demands that subsequent computational campaigns add all known Hh oxysterols as well as similar molecules to any analysis.

With regards the virtual screens carried out using molecules from the ZINC dataset, the hydrogen-bonding network in the middle pocket (R400, H470, D473, E518) provided, *prima facie*, several tempting polar residues that have been targeted by ligands with larger moieties in this position and, by itself, is believed to stabilise SMO's inactive conformation [218].

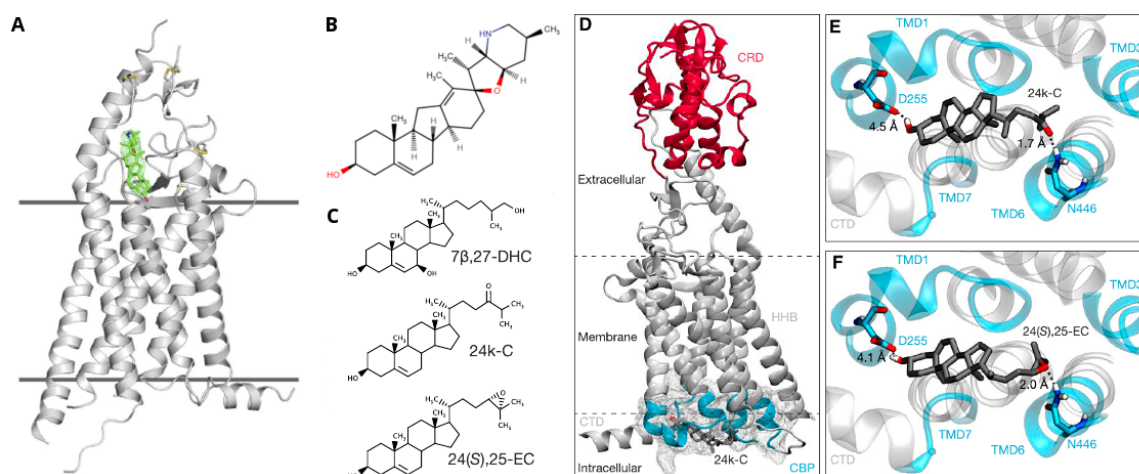


Figure 4.24: Other work involving oxysterol inhibition of smoothened. (A) The crystallographic binding position of Cyclopamine (green, stick representation surrounded by mesh) is displayed at the top of the heptahelical bundle near the entrance to the binding pocket, an example of an potent oxysterol-like ligand perhaps too bulky to bind deeper in the TM7 bundle of SMO (from Weierstall, et al, 2014). (B) 2D structure of Cyclopamine showing its largely apolar structure (C) 2D structures of various oxysterols from Raleigh, et al, 2018, depicting a bulky fused-ring segment with short aliphatic tail sections. (D) Depiction of SMO crystal structure with its CRD (red cartoon), C-Terminal Domain (CTD, grey cartoon) and Cytoplasmic Binding Pocket (CBP, aqua cartoon) as defined by the authors of Raleigh, et al, 2018. (E-F) CBP for SMO showing key CBP residues (aqua cartoon) and predicted binding mode of two oxysterols (grey stick representation) from Rayleigh. et al, 2018.

However, the search for chemical novelty was not the only goal as some parameters for useful SMO ligands needed to be met. In accord with one of the stated goals of both virtual screens, all tested compounds were predicted to avoid D473, a common mutation in SMO. Structural rearrangements of the SMO binding pocket from mutations such as D473G, D473N and D473H have been demonstrated to abolish binding of Vismodegib entirely [199] and significantly reduce the affinity of other inhibitors such as AntaXV [198]. Other residues prone to mutation (e.g. E481) were avoided.

Even with these restrictions, cellular assays confirmed hits in both rounds of ZINC docking with hit rates of 25% and 75% respectively, hit rates somewhat higher than other SMO campaigns but within the usual range of many GPCR virtual screens (17% - 58% [212]). This was achieved whilst specifically avoiding similarity with previous ligands, residues prone to mutation and also despite a well known bias in commercial compound libraries toward well-studied targets [219] [220] (e.g. β 2 Adrenergic Receptor). Both of these factors severely limit the pool of inhibitors that may bind similarly to known SMO ligands.

Whilst previously known ligands and their binding modes provided a guide for choosing possible inhibitors, the focus was most certainly on finding novel binding modes. That is, candidate inhibitors predicted to interact in a similar manner to known SMO ligands were

only prioritised if they displayed interactions in addition to those demonstrated by known SMO ligands. This is simply because SMO ligands tend to be driven by apolar binding (e.g. Cyclopamine) and then only one or two polar interactions within the same loop or helix. For example, Vismodegib only displays polar interactions with residues in ECL3 (D384 and Y394) and otherwise exhibits apolar binding. A more potent ligand, SANT-1, again we see only two polar interactions but with residues in both ECL3 (Y394) and TM6 (H470). Accordingly, candidate inhibitors displaying more electrostatic interactions were naturally prioritised. As were those where molecules were predicted to interact with multiple loops and helices.

Specific attempts to discover new 'deep' binding ligands appear to have borne fruit. As SANT-1 is the only known ligand to bind in this way, and is so affine it was used as the benchmark inhibitor in all cellular assays conducted for this work, it was considered important to find more molecules that bind as deep as SANT-1 and again prioritise novelty. All molecules tested that were predicted to be deep binders showed concentration-dependent decreases of constitutive signal activity in the cellular assay. All were predicted to interact with similar residues (Y394) and penetrate deep into the pocket below H470 but priority was also given to molecules predicted to interact with residues higher in the pocket such as those in the hydrogen-bonding network. Those also interacting with H470, as in SANT-1 displayed a greater decrease in residual Hh signal on the cellular assay. That the hit rate significantly increased in the second screening for the relaxed Vismodegib docking structure also confirms the importance of shape in docking.

With this in mind, two emerging Structure-Activity Relationships can therefore be elucidated from assay results. In Round 1, similar Hh signal reductions were observed for **K5** and **K6** at 100 μ M with **K5** significantly better at 10 μ M. Both molecules possess charged cyclic amines predicted to form a salt-bridge with the carboxylate sidechain of D384 in ECL1 and amide nitrogen predicted to interact with E518 in TM6. However, **K5** also had an additional predicted interaction between Y394 in ECL3 and its pyridine, also observed with Vismodegib (Figure 4.25). Whether the additional interactions with E518 result in a more affine ligand is to be determined in a binding assay.

A second trend was observed in **K2.7**, **K2.11** and **K2.12**. All three molecules possess a β -hydroxy moiety predicted to interact with various middle-pocket residues (Figure 4.26). However, assay results showed greater Hh signal reduction for **K2.7** and **K2.12**, both of which are predicted to link trans-membrane helices and extra-cellular loops via salt-bridge

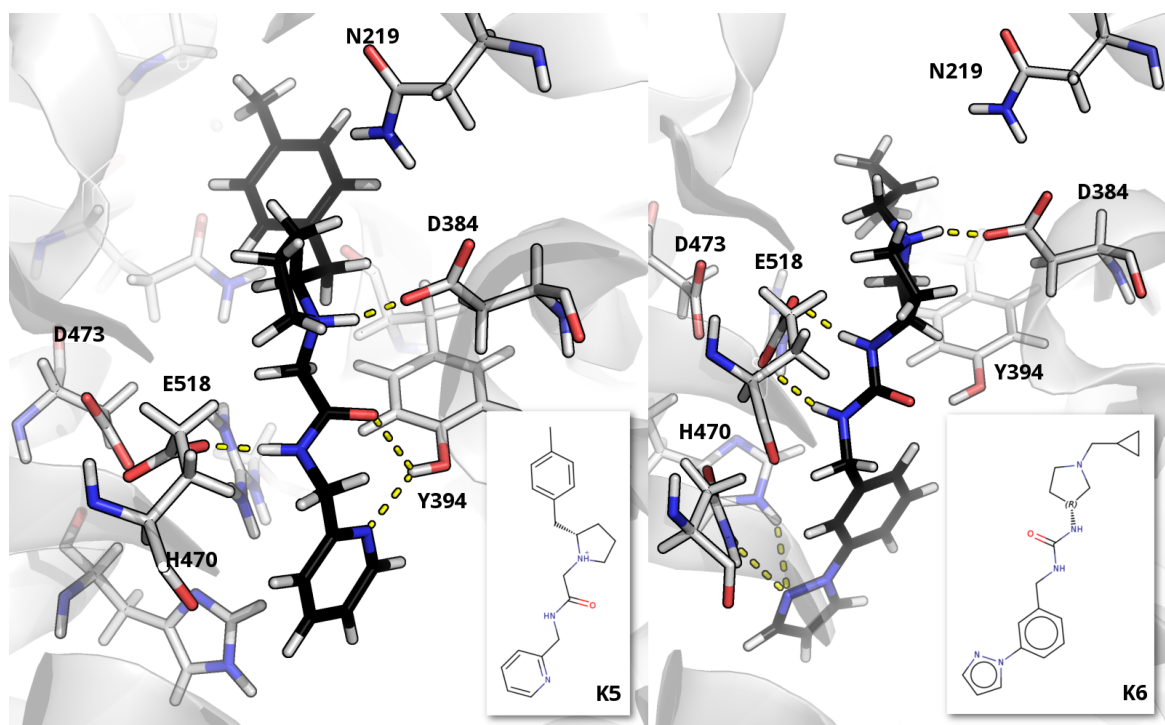


Figure 4.25: Predicted binding modes of **K5** (left, black carbons, stick representation) and **K6** (right, black carbons, stick representation) in the SMO binding pocket (TM7 bundle). Key residues (light grey, stick representation) are labelled, predicted polar interactions are represented by yellow dashed lines. In each case, predicted salt-bridge interactions with ECL3 residues (D384) are complemented with a polar interaction between an amide/urea Nitrogen and a residue in TM6 (E518). As per assay results, **K5**'s greater ability to impede Hh signalling at a lower concentration (10 μ M vs 100 μ M) is suggestive of the additional interaction with Y394 being more important in terms of Hh signal reduction than H470.

interactions with E518 in TM7, Y394 in ECL3 with respect to **K2.7** and D384 in ECL1 with respect **K2.12**.

It is noteworthy that **K2.7** reduces Hh signalling to the level of SANT-1. This is despite no optimisation ligands from either ZINC screen. Despite uncovering some ligands active against SMO that were predicted to bind deeper in the binding pocket, none appeared to reduce Hh signalling as much as SANT-1. Binding assays will be required to confirm whether **K2.7** is as affine as SANT-1 and efforts should be made to test for colloidal aggregation [164], a common but difficult to predict problem for any inhibitor development campaign.

However, more molecules were found to be active against SMO in the second round of testing that were also predicted to interact with similar residues to **K2.7**. This suggests that these residues (Y394, E518) are likely more accessible and opens possibilities for optimisation of both interactions and chemotypes without the need to bind deeper and to compete with SANT-1 on affinity.

More generally, what structural features make a good SMO ligand? In terms of the binding

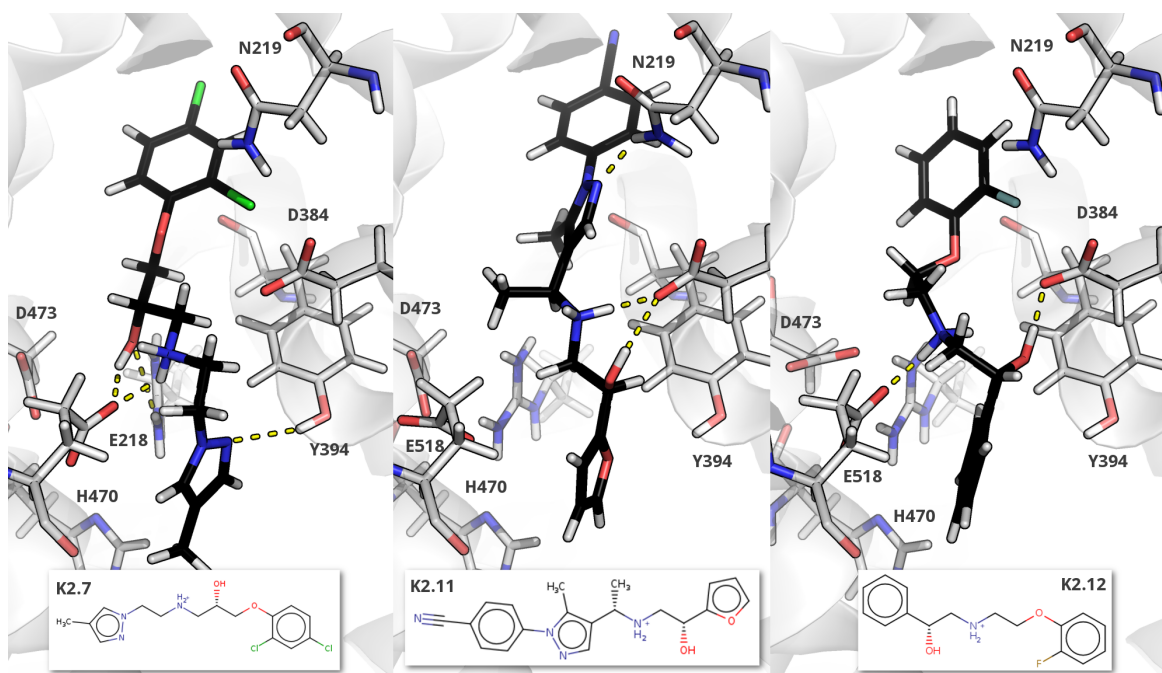


Figure 4.26: Predicted binding mode of **K2.7** (left, black carbons, stick representation), **K2.11** (middle, black carbons, stick representation) and **K2.12** (right, black carbons, stick representation) in the SMO binding pocket (TM7 bundle). Key residues (light grey, stick representation) are labelled, predicted polar interactions are represented by yellow dashed lines. As per assay results, ligand **K2.7** displayed greater Hh signal reduction at 100 μ M than both **K2.11** and **K2.12**, comparable to the positive control (SANT-1). **K2.12** displayed greater Hh signal reduction than **K2.11** at the lower concentration (10 μ M) and, like **K2.7**, was also predicted to link residues in transmembrane (E518) and extra-cellular helices (D384). These predictions underscore the likely importance of Y394 in SMO ligand binding and supports the strategy to select for molecules that link transmembrane and extra-cellular helices.

pocket, Y394 in ECL3, in isolation, appears to be a crucial residue for ligands binding to SMO's inactive conformation. The pyridine nitrogen of Vismodegib interacts with the hydroxy sidechain of Y394 and its binding within the TM7 bundle is associated with a conformational change of ECL3 that occludes the sterol binding groove in the CRD occupied by active-state SMO [201]. Similarly, a pyrazole in the SANT-1 X-ray structure also interacts with Y394 [198]. Results from signalling assays performed support this, molecules possessing predicted interactions with Y394, regardless of the chemotype, generally result in a decreased Hh signal activity (Figure 4.27).

Secondarily, molecules displaying predicted interactions bridging Y394 and residues in TM6 such as E518 were associated with greater decreases in Hh signalling. A very recent structure optimisation campaign of a single virtual screening hit combining competitive binding assays and free-energy methods (MM/GBSA) confirmed the primacy of the binding free-energy contribution of Y394 but there were large gains in affinity from also targeting E518, whose energy contribution was only slightly less [221]. Structurally, amide or urea moieties appear ideal to exploit this linking of ECL3 and TM6 as the rigidity and geometry of these functional

groups are easily within reach of the hydroxy sidechain of Y394 and the carboxylate of E518, particularly in the case of compounds such as **K5**, **K11** and **K2.19** (Figure 4.27).

Additionally, molecules displaying predicted interactions with another residue in TM6 (H470) in concert with either or both of Y394 and E518 were also associated with large decreases in Hh signalling, particularly **K6** and **K2.19** (Figure 4.27). This effect was not consistent, however, and of the three residues discussed here, despite as potent a ligand as SANT-1 also displaying this interaction with SMO, it may be the least important of the three residues discussed here.

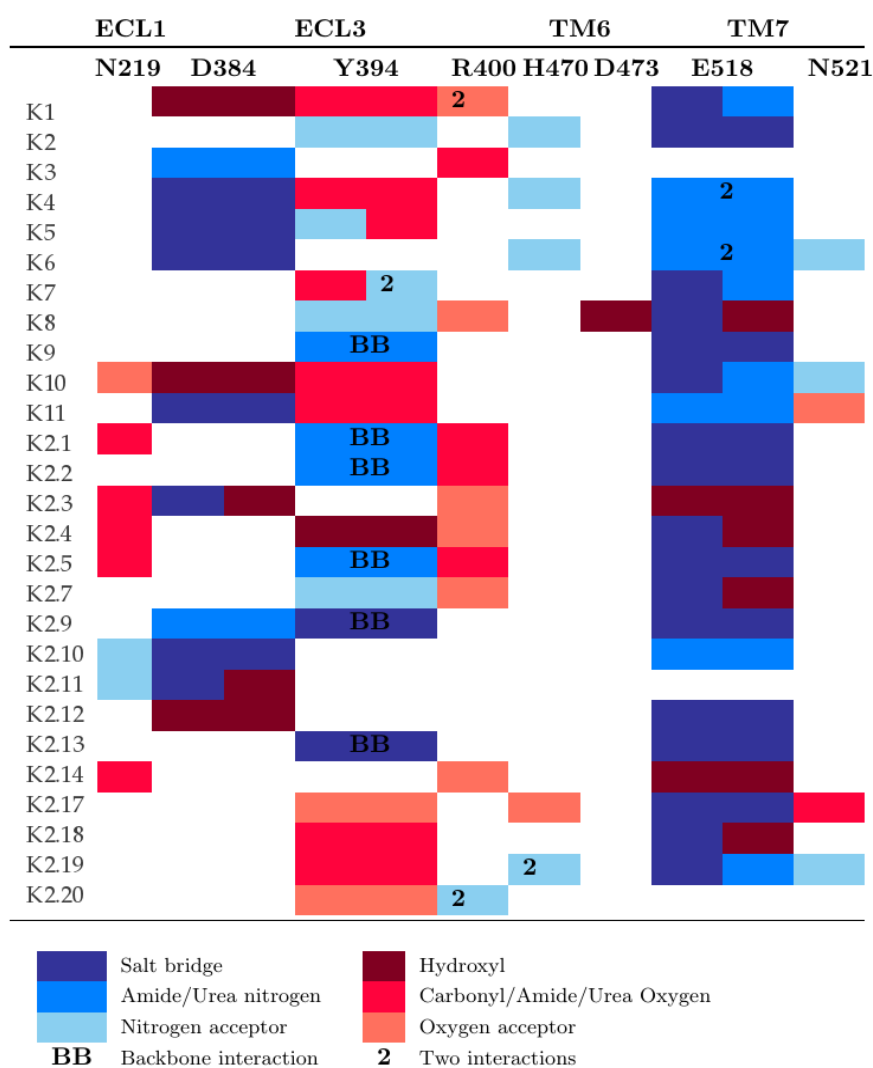


Figure 4.27: Interaction matrix of all compounds tested in Rounds 1 and 2. Colours represent different interacting moieties in line with traditional colours of atoms in chemistry (red=Oxygen, blue=Nitrogen). Generally, compounds predicted to interact with Y394 displayed a greater propensity to reduce Hh signalling, followed by those predicted to link residues in ECL loops and TM helices (e.g. Y394 and E518, respectively). Compounds interacting with H470 also displayed a greater propensity to reduce Hh signalling but the effect was not consistent, hence Y394 and the strategy of linking ECL loops and transmembrane helices (e.g. Y394/D384 and E518) appear to be more important for binding to SMO's TM7 bundle.

Tarting of residues in the SMO binding pocket to encourage ‘deep’ binding was less effective than pocket minimisation in retrieving hits. In principle, tarting a lot of residues should dramatically increase overall electrostatic potential surrounding affected residues and highly polar molecules should therefore score higher. This was borne out as more highly polar decoys outscored actives in the tarted receptor in initial tests. However, docking scores for less polar actives against SMO were largely unaffected by tarting. Several known actives, poorly scored in the untarted receptor, did however score highly in the tarted receptor.

It was therefore possible that more molecules active against SMO, polar but not *too* polar, could be retrieved. The increased risk of retrieving molecules from ZINC that score high purely because they are more polar was mitigated as molecules in the ZINC dataset tend not to be as apparently polar as generated decoys. This was confirmed as docked ZINC12 molecules tended to be scored and ranked similarly in both tarted and untarted receptors. A few known active molecules were nonetheless retrieved against the tarted receptor after being missed in docking to the untarted receptor (e.g. **K2.19**).

A few facets of the SMO binding pocket allow for a greater chance for success in any structure-based screen of this nature. As with other GPCRs, SMO has a deep and narrow cavity in its transmembrane domain and is generally not accessible by solvent. It is larger than the orthosteric binding pocket of aminergic receptors such as the β 2AR but significantly smaller than that of peptide receptors and shape-complementarity is important for docking. The library bias of commercially-available libraries such as ZINC toward GPCRs increases the probability of retrieving molecules that meet well with the shape-complementarity desired in SMO ligands, reflected in the higher than usual hit-rate. Underscoring this, as mentioned, **K2.7**, **K2.11** and **K2.12**, possess a β -hydroxy moiety that has been proven important for binding to specific residues in the β 2AR.

Thus, from a structure-based screen of libraries biased toward availability, 15 ligands with little commonality between their structures were retrieved and their antagonism to SMO conformed via cellular assay. The importance of a structure-based perspective was demonstrated when relaxation of the binding pocket prior to docking found that small changes in the positions of residues around the binding position of Vismodegib resulted in the retrieval of more antagonists. The importance of key extra-cellular and transmembrane residues was highlighted in addition to the enhanced activity of compounds that target multiple residues across domains.

The compounds themselves were demonstrated to be active with predicted binding modes

that avoid residues prone to mutation and consequent clinical resistance and at least one was potentially as potent as the most affine Hh pathway inhibitor known. Additionally, without optimisation, several drug-like compounds displayed significant activity against SMO with chemotypes that are, in principle, easily synthesisable and amenable to optimisation. All of this contributes to structural knowledge of SMO and provides excellent starting points for further structure-based campaigns for the development of more potent inhibitors.

Chapter 5

Perspective

5.1 Introduction

The greatest costs of any fragment-based campaign are those resulting from an in-depth analysis of false positives whereas false negatives are a total loss of important information. Using computational methods in concert with experimental validation, a campaign of ligand discovery can reduce the probability of such losses. The previous chapters describe specific attempts to find small molecule inhibitors against two *prima facie* straightforward molecular targets, Pim-1 and SMO. As with anything in life and research, the devil was in the tiny but significant details.

Chapter 4 focussed on the development of both synthetically feasible and optimisable ligands against Pim-1. Along the way, insights regarding the dynamics and structure of the Pim-1 binding pocket were gained via MD simulation and a docking comparison that informed subsequent structure-based virtual screening campaigns. Although synthetic feasibility remains a difficult problem, some success was observed in synthesising inhibitors that were active against Pim-1 as well as some amenable to further optimisation from a later screen using the ZINC dataset. Some surprising results involving a stabilising peptide are suggestive of future development of less-explored sections of the Pim-1 binding pocket and, with some modifications, a crystallised fragment offers attractive extensions for future Pim-1 inhibitors.

Ligands discovered in chapter 4 were both chemically novel and active against SMO. These ligands were predicted to make interactions distinct from previous SMO ligands within the allosteric pocket of SMO in the 7TM bundle and were chosen specifically to avoid known undesirable effects (e.g. interactions with D473, known to be prone to mis-sense mutation). Minimisation of the SMO binding pocket generally retrieved more ligands and emerging structure-activity relationships gave additional insight into important residues for future ligand discovery campaigns.

In light of the above, it bears examining in detail what was learned, both positive and negative, from the multifaceted investigations of these two targets and how these lessons can be applied to future work involving these two targets.

5.2 What was learned from this thesis

5.2.1 Chemically diverse synthetically feasible ligands were found against Pim-1

A campaign was described in chapter 3 that focussed on the discovery of novelty and synthetic feasibility in ligands for the Pim-1 kinase. A number of small ($< 250\text{Da}$) molecules from a diverse in-house library were predicted and confirmed to bind via a docking and crystallographic screen. These fragments, in and of themselves, are unsuitable as a drug due to the nature of their unspecific binding and generally have low affinity for any target. However, as demonstrated, fragments with reliable binding modes make an ideal starting point for extensions into other sections of the Pim-1 binding pocket.

Although not novel *per se*, the diverse nature of fragments from the in-house library and their known binding modes enabled consideration of productive extensions. A focus on synthetic feasibility led to enumeration of points on fragments amenable to extension and the use of a published set of ‘safe’ synthetic reactions. An additional criterion of availability determined that these molecules would be extended via purchasable ‘building blocks’ from the ChemBridge library.

Although the focus on synthesisability and availability potentially compromised the degree of chemical space exploration in this project, there was some level of compensation for loss of novelty as specific sections of the Pim-1 binding pocket could be targeted by molecules synthesised to do so. Coupled with a reliable fragment binding mode, this enabled some risks to be taken. There are few ligands targeting those near Pim-1’s Specificity Surface (e.g. D128, D131, E171) and molecular choices during evaluation of docking results guided us toward making these residues a priority. This is despite the risks inherent in targeting residues that are both somewhat divergent from ideal shape complementarity and solvent-exposed.

The result of this campaign was some ligands with good enough affinity to form the basis for optimisation, traditional in screening campaigns against kinases. These were predicted to address seldom targeted parts of Pim-1 and were inactive against a selectivity target, Protein Kinase A. It was found that molecules showing activity were generally more rigid than other known Pim-1 inhibitors, suggestive of an important facet of any molecule targeting residues near the Specificity Surface.

An additional factor appears to have been the nature of the electrostatic interaction with those residues. Measured activity was highest with ligands that had uncharged moieties interacting with solvent-exposed charged residues. This is suggestive that solvent screening ‘dampening’ effects were less of a problem than with charged moieties. This was certainly the case with the most potent compound, a hydroxy benzofuran-benzoic acid, where its hydroxy group was predicted to interact with D128 and its carboxylate predicted to interact with Pim-1’s catalytic lysine (K67). Compounds possessing a benzoic acid but coupled to a charged moiety (e.g. piperidine) were inactive in assays. The effect seems to be ameliorated somewhat by substitution of the charged species with, for example, a short aliphatic chain. However, the activity of one compound with an additional group of this nature was still relatively weak with a three-digit K_i .

5.2.2 Novel ligands were also found in focussed libraries for diverse targets

Project 3 in chapter 3 and two ZINC screens in chapter 4 also unearthed chemical novelty for both Pim-1 and SMO. This is perhaps a slightly surprising result as the ZINC dataset is perhaps less likely to retrieve kinase ligands. However, several potent hits were observed against SMO with room for further development and optimisation.

With regards Pim-1, proposed inhibitors were tested via orthogonal assays, some were crystallised and X-ray structures obtained. Those showing activity were predicted to interact with Pim-1 via canonical binding modes with the hinge and/or catalytic lysine (K67). However, chemically they were novel when compared to other Pim-1 ligands, generally comprised of more flexible linkers rather than the large apolar moieties traditionally associated with Pim-1 inhibitors. The most potent, however, was also the most rigid, a triazolo-thiadiazole coupled to an indole on one side and a pyridine on the other.

Two rounds of a ZINC screen were also successful in finding ligands for SMO. In some ways this is not a surprise as the commercial compound libraries it is based on have a well-known bias toward GPCRs. However, hit rates in both rounds were above expectations and results showcased the best of a structure-based virtual screen. Molecules showing reasonable activity were generally dissimilar to previously published ligands and patterns of activity in ligands with a degree of similarity to each other suggested at least one emerging Structure-Activity Relationship (SAR). Additionally, ligands were found in assays that lowered signalling activities to within range of the most potent inhibitor known against SMO.

So for both targets, chemically diverse hits have been the result of screens of the ZINC library.

None of the ligands appeared potent in drug terms but the nature of the subset used, drug-like and filtered for availability, has left some ‘headroom’ for a productive optimisation campaign. This is the case in terms of the ligands themselves as well as for similar compounds.

5.2.3 Structural information matters

In Project 1 (chapter 3), the importance of binding pocket shape in proteins was highlighted. The project was designed to address a problem with any virtual screening campaign, that of making the choice of a starting structure for docking. Related, as fragments often find multiple poses in a binding pocket, the problem of finding a reliable binding mode to extend in fragment-based approaches as well as choosing an appropriate docking program was analysed.

Starting structure choice often requires expert input, particularly as subtle differences in conformation can affect the chemotypes predicted to bind in an active site as well as the quality of those predictions. A work flow including MD simulation of a bound ligand-structure coupled with docking using multiple software codes was demonstrated to improve performance across the board in ‘cognate’ docking, or, the ability of a docking program to retrieve a ligand pose close to the known X-ray crystal structure. This is presumably as the simulations were able to explore conformational states that reflect a ‘typical’ bound ligand.

A conformation was found for subsequent docking that allowed for the retrieval of a more diverse array of ligands than is reflected in the structure of the original bound ligand. It may in fact be the case that the particular bound ligand matters less than simulation with *any* given ligand. By way of contrast, in the absence of a bound ligand, simulation from a bound ligand state generally worsened docking performance as conformational states reflecting the unbound state of the protein were sampled. The overall improvement in docking performance of ligand-bound simulated structures, therefore, both obviates the need for expert choice and reliance upon the choice of software.

Minimisation of the SMO binding pocket also demonstrated the importance of conformation for rigid docking as, despite docking to a drug-bound structure, initial results were less than satisfactory. Even after ligand minimisation, in some cases poses displayed extreme torsional strain. There was also an unusual frequency of very unlikely protomers placed within the pocket that scored highly - indicative of a sampling problem. This led to the conclusion that the binding pocket itself may not be in a conformation likely to retrieve more novel ligands and with reasonable poses. Minimisation of residues around the position of the ligand in the

drug-bound structure quite simply led to more ligands that were reflective of a reasonable binding position.

The provision of exhaustive ligand conformations mattered less than protein structure in all projects. The shape of the molecular pose is crucial in prediction of the most likely binding mode in biomolecular studies. However, particularly with the fragment study of Pim-1 in Project 1 of chapter 3, only a few conformers (5-15) in energetically likely conformations were necessary to retrieve poses close ($< 2\text{\AA}$) to the crystal binding mode for some fragments. Only a few more conformers per molecule (15-20) were generated in libraries of extended molecules in Project 2 and Project 3 and for both ZINC screens against SMO in chapter 4.

The number of rotatable bonds in molecules generated by the DOCK pipeline also affects the number of conformers generated. An ‘energy window’ parameter in conformation generation dictates accept/reject criteria for conformers. If the difference in molecular strain energy between a conformer and the global minimum conformer is above a certain threshold, the conformer is omitted from preparation. This parameter is set progressively much more strictly for molecules with a higher numbers of rotatable bonds, limiting the number of conformers generated and, ostensibly, reducing computational time per molecule. This, of course, has the effect of limiting the amount of chemical space explored by rigid fragments supplied for docking.

However, it would appear that the lack of conformational space explored by fewer conformers is compensated for the fact that all docking codes used have a minimisation step after initial docking of the rigid molecule. This essentially kicks retrieval of the most likely conformation for each pose down the road. Energy minimisation will not resolve large clashes with the protein surface or eject the molecule from the binding pocket. However, it can certainly change pose geometry enough to make and break interactions or fatally affect shape complementarity in correcting the geometry of, for example, amines to reflect sp^2 -hybridisation rather than their putative sp^3 geometry. The implicit assumption is therefore that it is more important that an initial docked position of a rigid molecule be converted to a favourable pose after minimisation than the generation of more conformers prior to docking.

5.2.4 One should always be prepared for surprises

Two unexpected results were observed in this thesis. Firstly, in Project 3 of chapter 3, two ligands were crystallised in complex with Pim-1 and a consensus peptide, Pimtide. Pimtide is a consensus peptide that was designed to mimic the binding of the Pim-1 co-substrate and

is also used to stabilise the protein-ligand complex for crystallography. Secondly, in Project 4 of chapter 3, a fragment from the internal fragment library, f200, was crystallised with only its quinoxaline moiety, missing entirely a thiophene and unsaturated linker. As the original structure for this fragment had these functional groups present, this fragment was initially missed when docking the internal fragment library from where it had originated.

Both ligands in Project 3 of chapter 3 crystallised and displaying water-mediated interactions with Pimtide. Both of these results are the first examples of direct interactions between a ligand and Pimtide on the PDB, albeit in only one ligand is this directly observed in the X-ray crystal structure. As the initial docking was completed without Pimtide in the model, the predicted pose was very different to the eventual binding mode. In fact, the only way to retrieve the crystal binding mode was to dock the ligand in a protein prepared with Pimtide and all crystallographic waters present. Shape complementarity was, again, demonstrated as being the key influence determining ideal binding positions in protein-ligand complexes.

Residues near the Pim-1 Specificity Surface are presumably difficult to target due to their degree of solvent exposure. The position of both crystallised ligands means they are in complex with a crystallographic water, D128 and R6 on Pimtide. A goal for the development of various peptide substrates has been to analyse protein kinase phosphorylation specificity but residues nearby to Pimtide's usual binding position, less conserved in kinases, offer the possibility other peptide inhibitors. This should be aggressively pursued as the lack of DFG-loop dynamics means active-state Pim-1 has been extensively targeted so finding chemical novelty in the small-molecule space will be progressively more difficult. Targeting of these conserved residues, of course, offers the possibility of more selective Pim-1 ligands too.

The truncated form of the fragment ligand in Project 4 of chapter 3 was initially missed as the full fragment was that which was used in docking. Attempts to find a docked position for the full fragment from the position of the X-ray structure inevitably resulted in clashes with the protein surface and attempts by in-house specialists to model the full fragment into measured electron density were not able to supplant evidence in favour of the truncated fragment. The X-ray crystallographic structure of the truncated ligand was, however, suggestive of an ideal starting structure for a fragment extension project. Apparent interactions between the amide carbonyl of its quinoxaline core and Pim-1's catalytic Lysine (K67) as well as the amide nitrogen with a buried water suggested a stable binding mode. The position of the apolar part of the fragment in within Pim-1's hydrophobic pocket should only have strengthened its position.

Sadly, it was found that neither the quinoxaline core nor several extensions tested were active against Pim-1 in a FRET-based assay. The most likely explanation seems to be that, although extended molecules were chosen with polar interactions in mind, none of the extensions approached the hinge. Many Pim-1 inhibitors have large apolar moieties within range of the hinge, naturally more apolar in Pim-1. A remedy along these lines would be to target the hinge by extending the quinoxaline fragment at a different position, for example 7 instead of 6. This has the benefit of concomitantly optimising for shape complementarity of the Pim-1 binding pocket.

5.2.5 Some aspects didn't work as well as we would like

Some projects were left with somewhat unsatisfactory results. Specifically, that there were few ligands found in Project 2 of chapter 3 against Pim-1, that the goal of synthetic feasibility remains a difficult one and that the CURED found in the SMO allosteric binding pocket in chapter 4 was unable to be identified.

Lack of affinity caused by the possibility of solvent screening effects between charged ligands and solvent-exposed charged residues in Project 2 was probably exacerbated by the flexibility of molecules that were found inactive against Pim-1. Most potential inhibitors had relatively flexible linkers between moieties within the binding pocket and those predicted to interact with residues such as D128.

Docking, particularly rigid docking, typically does not attempt to quantify configurational entropy changes upon ligand binding [222]. Some attempts using rigid multiple poses have been reported [223]. But, in and of themselves, docking scoring functions are unable to quantify the degree to which these effects, that oppose the attractive forces involved in binding [224], will impact on docking predictions.

Appropriate sampling of conformational space via molecular simulation coupled with free energy methods (e.g. Free Energy Perturbation) are designed specifically to estimate and decompose the various free energy changes upon ligand binding. These methods do not often achieve so-called 'chemical accuracy' ($< 1 \text{ kcal}\cdot\text{mol}^{-1}$), ostensibly for reasons related to inadequate sampling, leading to Mean Absolute Deviations (MAD) from experimental estimates more of the order of 4-6 $\text{kcal}\cdot\text{mol}^{-1}$ [225]. However, they can be useful for qualitative ranking of molecules in a series of similar ligands, for example. Addressing the problem of solvent polarisation effects is more tricky. Again, for qualitative ranking, free energy methods using polarisable force fields (e.g. AMOEBA [226]) or even hybrid methods (e.g. QM/MM) should

be useful in congeneric ligand series'. As these are computationally-expensive methods at present, the aim as hardware speeds increase will be to scale up their use to broader sets of ligands.

Synthetic feasibility was the goal in the development of ligands in Project 2 of chapter 3 and results were quite mixed. In almost all cases, compounds were synthesised with methods that differed to that proposed by our software. In many cases, molecules proposed were required to be altered slightly to make them synthesisable either at all or not without significant effort and development expended by synthesis collaborators.

There is no easy automated solution for this. Whilst some syntheses do prove to be relatively straightforward, *in silico* reactions should remain more as idea generators than a blueprint for automated synthetic success. It therefore remains that even apparently straightforward organic molecule synthesis requires expert intervention for all synthesis methods.

Identification of the molecule(s) within the electron density in chapter 4 also proved elusive. The endogenous ligand of SMO is, at present, not known and many compounds used in SMO's crystallisation in addition to assay kit ingredients were ruled out of consideration during X-ray model refinement. Docking of databases including compounds known to be present in cell process such as membrane transport and signal transduction as well as cell metabolites failed to find a candidate ligand. Docking of small molecular probes provided valuable insight into the energetics of various segments of the SMO binding pocket and did allow us to narrow the list of moieties likely to be favourable in these segments. However, in terms of raw combinatorics, a library of molecules including all possibilities of interacting moieties and appropriate linkers would be huge.

A proposed approach is to build a library of molecules based on filtering rules for specific segments of the CURED then docking them. From there, the application of, for example, a scoring function based on spatial criteria [227] could be applied to score and rank the possibilities. One imagines that would at least further narrow molecular possibilities but a run-on effect is the possibility of identifying other molecules that could plausibly fit within the binding pocket to discover new ligands for SMO.

5.2.6 What is a 'hit'?

Results from both major targets in this thesis prompt the question of what constitutes a real 'hit' in ligand discovery campaigns. Does it have pM affinity for the target? Or merely

those that show enough to be interesting for further development? Are these criteria target-dependent? If a compound shows activity but crystallises in an entirely different position than that predicted, does this no longer count as a hit? And does the limited nature of biophysical methods mean that we cannot be sure of a molecule's status as a 'hit' until we obtain an X-ray structure?

In terms of raw affinity, the answer to this question is a resounding no for all ligands tested. None achieved even nM affinity so in kinase terms, the best ligands at two-digit μM affinity would be considered relatively weak ligands. However, as stated, most kinase campaigns begin with ligands inhibiting at this level of affinity. This is closely followed by an intensifying of effort in the service of optimisation that can take some months or even years prior to testing (e.g. *in vivo* assays) and publication.

Sometimes, despite the success of a crystal hit, it may not be considered a *real* hit. As reported in Project 3 of chapter 3, two molecules displayed completely different binding modes to that predicted by docking, due to the unforeseen intervention of Pim-1's consensus peptide Pimtide. Unfortunately, one of those crystallised did not show activity in an assay. A third ligand was shown to be active in assays and was crystallised but, although a dataset was obtained, it did not survive structural model refinement as the ligand could not be accurately placed within the electron density measured.

It would be tempting to view those interacting with Pimtide as not real hits. However, although not reflective of the exact sequence of Pim-1's co-substrate, one of the aims of Pimtide is to mimic substrate binding. Additionally, the positions of both ligands should inhibit ATP binding. It is therefore possible that the binding mode observed with these ligands reflects a bound state where ATP is replaced but allow for co-substrate binding. This was not the case where one ligand was crystallised with Pimtide but inactive in assays. However, it does not preclude others from this possibility.

One of the ligands crystallised with Pimtide showed activity in assays. Assuming it was not a hydrolysis product that crystallised, this is suggestive of more than one binding mode. The binding mode elucidated from its X-ray structure still counts as a X-ray crystallographic hit and the information gained from its binding position will be useful in targeting less conserved Pim-1 residues.

A third example, whilst it did not progress to a fully refined structure, was active both against Pim-1 but also showed signs of pan-PIM inhibition. Primarily, the limitations in ligand

detection using biophysical methods concern false negatives so despite no X-ray structure and a relatively weak affinity, its status as a hit should be assured.

To summarise, the determination of a molecular ‘hit’ depends upon the criteria used. Some compounds discovered that make a reasonable starting point for optimisation or show signs of other useful properties would be disregarded as hits by others. What can be gleaned from this thesis, however, is that although there are costs associated with pursuing leads that perhaps end less than satisfactorily, the costs of false negatives are known to be much greater. So care should be taken before outright rejecting compounds that show reasonable signs of activity but have scope for optimisation or tell us something about a target that was not previous known.

Chapter 6

Bibliography

-
- [1] L. M. Mayr and P. Fuerst, "The future of high-throughput screening," *J. Biomolecular Screening*, vol. 14, no. 16, pp. 1923–1938, 2014.
- [2] R. Liu, X. Li, and K. S. Lam, "Combinatorial chemistry in drug discovery," *Curr. Opin. Chem. Biol.*, vol. 38, pp. 117–126, 2017.
- [3] A. Carnero, "High throughput screening in drug discovery," *Clin. Transl. Oncol.*, vol. 8, no. 7, pp. 482–490, 2006.
- [4] T. D. Pollard, "A guide to simple and informative binding assays," *Molecular Biology of the Cell*, vol. 21, pp. 4061–4067, 2010.
- [5] M. E. Welsch, S. A. Snyder, and B. R. Stockwell, "Privileged scaffolds for library design and drug discovery," *Curr. Opin. Chem. Biol.*, vol. 14, no. 3, pp. 347–361, 2010.
- [6] B. Raymer and S. K. Bhattacharya, "Lead-like drugs: A perspective," *J. Med. Chem.*, vol. 61, no. 23, pp. 10375–10384, 2018.
- [7] J. Wang and T. Hou, "Drug and drug candidate building block analysis," *J. Chem. Inf. Model.*, vol. 50, no. 1, pp. 55–67, 2010.
- [8] S. G., "Prediction of drug-like properties," *Madame Curie Bioscience Database*, 2000.
- [9] B. K. Shoichet, "Virtual screening of chemical libraries," *Nature*, vol. 432, pp. 862–865, 2004.
- [10] A. C. Good and T. I. Oprea, "Optimization of camd techniques 3. virtual screening enrichment studies: a help or hindrance in tool selection?," *J. Comput.-Aided Mol. Design*, vol. 22, pp. 169–178, 2008.
- [11] F. Kanehisa, M. Tanabe, Y. Sato, and K. Morishima, "Kegg: new perspectives on genomes, pathways, diseases and drugs," *Nucleic Acids Res.*, vol. 45, pp. D353–D361, 2017.
- [12] A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, and J. P. Overington, "ChEMBL: a large-scale bioactivity database for drug discovery," *Nucleic Acids Res.*, vol. 40, no. D1, pp. D1100–D1107, 2012.
- [13] S. Kim, P. Thiessen, E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B. Shoemaker, J. Wang, B. Yu, J. Zhang, and B. SH, "Pubchem substance and compound databases," *Nucleic Acids Res.*, vol. 44, no. D1, pp. D1202–D1213, 2015.
- [14] T. Stirling and J. Irwin, "Zinc 15 – ligand discovery for everyone," *J. Chem. Inf. Model.*, vol. 55, pp. 2324–2337, 2015.
- [15] G. Schneider, *De novo molecular design*. Wiley, 2014.
- [16] I. Kola and J. Landis, "Can the pharmaceutical industry reduce attrition rates?," *Nat. Rev. Drug. Discov.*, vol. 3, no. 8, pp. 711–716, 2004.
- [17] T. A. Halgren, "Identifying and characterizing binding sites and assessing druggability," *J. Chem. Inf. Model.*, vol. 49, pp. 2324–2337, 2009.
- [18] P. Schmidtke, V. Le Guilloux, J. Maupetit, and P. Tufféry, "Fpocket: online tools for protein ensemble pocket detection and tracking," *Nucleic acids research*, vol. 38, no. suppl 2, pp. W582–W589, 2010.
- [19] D. Alvarez-Garcia and X. Barril, "Molecular simulations with solvent competition quantify water displaceability and provide accurate interaction maps of protein binding sites," *J. Med. Chem.*, vol. 57, pp. 8530–8539, 2014.

-
- [20] I. Bahar, T. Lezon, A. Bakan, and I. H. Shrivastava, "Normal mode analysis of biomolecular structures: Functional mechanisms of membrane proteins," *Chem. Rev.*, vol. 110, no. 3, pp. 1463–1497, 2010.
- [21] C. N. Nguyen, T. K. Young, and M. K. Gilson, "Grid inhomogeneous solvation theory: Hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril," *J. Chem. Phys.*, vol. 137, no. 3, pp. 044101–1–044101–17, 2010.
- [22] H.-J. Böhm and M. Stahl, "The use of scoring functions in drug discovery applications," in *Reviews in Computational Chemistry* (K. B. Lipkowitz and D. B. Boyd, eds.), vol. 18, ch. 2, pp. 41–87, John Wiley and Sons, 2002.
- [23] D. H. Williams and B. Bardsley, "Estimating binding constants – the hydrophobic effect and cooperativity," *Perspectives in Drug Discovery and Design*, vol. 17, pp. 43–59, 1999.
- [24] R. G. Coleman, M. Carchia, T. Sterling, J. J. Irwin, and B. K. Shoichet, "Ligand pose and orientational sampling in molecular docking," *PLoS ONE*, vol. 8, p. e75992, 10 2013.
- [25] O. Trott and A. J. Olson, "Autodock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," *J. Comput. Chem.*, vol. 31, no. 2, pp. 455–461, 2010.
- [26] M. McGann, "Fred and hybrid docking performance on standardized datasets," *J. Comput.-Aided Mol. Design*, vol. 26, no. 8, pp. 897–906, 2012.
- [27] R. Friesner, J. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, S. Mee, J. K. Perry, D. E. Shaw, P. Franci, and P. S. Shenkin, "Glide: A new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy," *J. Med. Chem.*, vol. 47, pp. 1739–1749, 2004.
- [28] G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor, "Development and validation of a genetic algorithm for flexible docking," *J. Mol. Biol.*, vol. 267, pp. 727–748, 1997.
- [29] Z. Zhou, A. Felts, R. A. Friesner, and R. M. Levy, "Comparative performance of several flexible docking programs and scoring functions: Enrichment studies for a diverse set of pharmaceutically relevant targets," *J. Chem. Inf. Model.*, vol. 47, no. 4, pp. 1599–1608, 2007.
- [30] H. J. Bohm, "The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure.," *J. Comput.-Aided Mol. Design*, vol. 8, pp. 243–256, 1994.
- [31] R. Wang, L. Liu, L. Lai, and Y. Tang, "Score: A new empirical method for estimating the binding affinity of a protein-ligand complex," *J. Mol. Model.*, vol. 4, pp. 379–394, 1998.
- [32] M. D. Eldridge, C. W. Murray, T. R. Auton, G. V. Paolini, and R. P. Mee, "Empirical scoring functions: I. the development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes," *J. Comput.-Aided Mol. Design*, vol. 11, no. 5, pp. 425–445, 1997.
- [33] M. Rarey, B. Kramer, T. Lengauer, and G. Klebe, "A fast flexible docking method using an incremental construction algorithm," *J. Mol. Biol.*, vol. 261, pp. 470–489, 1996.
- [34] P. Ferrara, H. Gohlke, D. J. Price, G. Klebe, and C. L. Brooks, "Assessing scoring functions for protein - ligand interactions," *J. Med. Chem.*, vol. 47, pp. 3032–3047, 2004.
- [35] M. Zheng, B. Xiong, C. Luo, S. Li, X. Liu, Q. Shen, J. Li, W. Zhu, X. Luo, and H. Jiang, "Knowledge-based scoring functions in drug design: 3. a two-dimensional knowledge-based hydrogen-bonding potential for the prediction of protein-ligand interactions," *J. Chem. Inf. Model.*, vol. 51, no. 11, pp. 2994–3004, 2011.

- [36] G. Neudert and G. Klebe, "Dsx: a knowledge-based scoring function for the assessment of protein–ligand complexes," *J. Chem. Inf. Model.*, vol. 51, no. 10, pp. 2731–2745, 2011.
- [37] T. Cheng, X. Li, Y. Li, Z. Liu, and R. Wang, "Comparative assessment of scoring functions on a diverse test set," *J. Chem. Inf. Model.*, vol. 49, no. 4, pp. 1079–1093, 2009.
- [38] S. A. Hayik, R. Dunbrack, and K. M. Merz, "Mixed quantum mechanics/molecular mechanics scoring function to predict protein–ligand binding affinity," *J. Chem. Theory Comput.*, vol. 6, no. 10, pp. 3079–3091, 2010.
- [39] T. Zhou, D. Huang, and A. Caffisch, "Is quantum mechanics necessary for predicting binding free energy?," *J. Med. Chem.*, vol. 51, no. 14, pp. 4280–4288, 2008.
- [40] J. Liu and R. Wang, "Classification of current scoring functions," *J. Chem. Inf. Model.*, vol. 55, no. 3, pp. 475–482, 2015.
- [41] P. J. Ballester, A. Schreyer, and T. L. Blundell, "Does a more precise chemical description of protein–ligand complexes lead to more accurate prediction of binding affinity?," *J. Chem. Inf. Model.*, vol. 54, no. 3, pp. 944–955, 2014.
- [42] J. Gabel, J. Desaphy, and D. Rognan, "Beware of machine learning-based scoring functions - on the danger of developing black boxes," *J. Chem. Inf. Model.*, vol. 54, no. 10, pp. 2807–2815, 2014.
- [43] M. Baker, "Fragment-based lead discovery grows up," *Nat. Rev. Drug. Discov.*, vol. 12, pp. 5–7, 2013.
- [44] C. W. Murray and D. C. Rees, "The rise of fragment-based drug discovery," *Nat. Chem.*, vol. 1, pp. 187–192, 2009.
- [45] H. Köster, T. Craan, S. Brass, C. Herhaus, M. Zentgraf, L. Neumann, A. Heine, and G. Klebe, "A small nonrule of 3 compatible fragment library provides high hit rate of endothiapepsin crystal structures with various fragment chemotypes," *J. Med. Chem.*, vol. 54, 2011.
- [46] J. Schiebel, N. Radeva, S. G. Krimmer, X. Wang, M. Stieler, F. R. Ehrmann, K. Fu, A. Metz, F. U. Huschmann, M. S. Weiss, *et al.*, "Six biophysical screening methods miss a large proportion of crystallographically discovered fragment hits: A case study," *ACS Chem. Biol.*, 2016.
- [47] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, "Experimental and computational approaches to estimate solubility and permeability in drug discovery and developmental settings," *Advanced Drug Delivery Reviews*, vol. 46, pp. 3–26, 2001.
- [48] N. Radeva, J. Schiebel, X. Wang, S. G. Krimmer, K. Fu, M. Stieler, F. R. Ehrmann, A. Metz, T. Rickmeyer, M. Betz, J. Winquist, A. Y. Park, F. U. Huschmann, M. S. Weiss, U. Mueller, A. Heine, and G. Klebe, "Active site mapping of an aspartic protease by multiple fragment crystal structures: Versatile warheads to address a catalytic dyad," *J. Med. Chem.*, vol. 59, no. 21, pp. 9743–9759, 2016.
- [49] I.-J. Chen and R. E. Hubbard, "Lessons for fragment library design: analysis of output from multiple screening campaigns," *J. Comput.-Aided Mol. Design*, vol. 23, no. 8, pp. 603–620, 2009.
- [50] M. N. Schulz, J. Landström, K. Bright, and R. E. Hubbard, "Design of a fragment library that maximally represents available chemical space," *J. Comput.-Aided Mol. Design*, vol. 25, no. 7, pp. 611–620, 2011.
- [51] J. Schiebel, N. Radeva, H. Köster, A. Metz, T. Krotzky, M. Kuhnert, W. E. Diederich, A. Heine, L. Neumann, C. Atmanene, *et al.*, "One question, multiple answers: Biochemical and biophysical screening methods retrieve deviating fragment hit lists," *ChemMedChem*, vol. 10, no. 9, pp. 1511–1521, 2015.
- [52] O. Ichihara, J. Barker, R. J. Law, and M. Whittaker, "Compound design by fragment-linking," *Mol. Inf.*, vol. 30, pp. 298–306, 2011.

-
- [53] B. C. Doak, R. S. Norton, and M. J. Scanlon, "The ways and means of fragment-based drug design," *Pharmacology and Therapeutics*, vol. 167, pp. 28–37, 2016.
- [54] M. A. Burlingame, C. T. M. B. Tom, and A. R. Renslo, "Simple one-pot synthesis of disulfide fragments for use in disulfide-exchange screening," *ACS Comb. Sci.*, vol. 13, no. 3, pp. 205–208, 2011.
- [55] F. Chevillard and P. Kolb, "SCUBIDOO: A large yet screenable and easily searchable database of computationally created chemical compounds optimized toward high likelihood of synthetic tractability," *J. Chem. Inf. Model.*, vol. 55, no. 9, pp. 1824–1835, 2015.
- [56] Chembridge, "Building blocks." (accessed Sep 24, 2018).
- [57] M. Hartenfeller, M. Eberle, P. Meier, C. Nieto-Oberhuber, K.-H. Altmann, G. Schneider, E. Jacoby, and S. Renner, "A collection of robust organic synthesis reactions for in silico molecule design," *J. Chem. Inf. Model.*, vol. 51, no. 12, pp. 3093–3098, 2011.
- [58] F. Chevillard, H. Rimmer, C. Betti, E. Pardon, S. Ballet, N. van Hilten, J. Steyaert, W. E. Diederich, and P. Kolb, "Binding-site compatible fragment growing applied to the design of β_2 -adrenergic receptor ligands," *J. Med. Chem.*, vol. 61, no. 3, pp. 1118–1129, 2018.
- [59] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The protein data bank," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 235–242, 2000.
- [60] N. Majeux, M. Scarsi, J. Apostolakis, C. Ehrhardt, and A. Caffisch, "Exhaustive docking of molecular fragments with electrostatic solvation," *Proteins: Struct., Funct., Bioinf.*, vol. 37, pp. 88–105, 1999.
- [61] M. Scarsi, J. Apostolakis, and A. Caffisch, "Continuum electrostatic energies of macromolecules in aqueous solutions," *J. Phys. Chem. A*, vol. 101, pp. 8098–8106, 1997.
- [62] B. K. Shoichet, D. L. Bodian, and I. D. Kuntz, "Molecular docking using shape descriptors," *J. Comput. Chem.*, vol. 13, no. 3, pp. 380–397, 1991.
- [63] M. L. Connolly, "Analytical molecular surface calculation," *J. Appl. Cryst.*, vol. 16, pp. 548–558, 1983.
- [64] M. L. Connolly, "Solvent-accessible surfaces of proteins and nucleic acids," *Science*, vol. 221, no. 4612, pp. 709–713, 1983.
- [65] E. C. Meng, B. K. Shoichet, and I. D. Kuntz, "Automated docking with grid-based energy evaluation," *J. Comput. Chem.*, vol. 13, pp. 505–524, 1992.
- [66] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner, "A new force field for molecular mechanical simulation of nucleic acids and proteins," *J. Am. Chem. Soc.*, vol. 106, no. 3, pp. 765–784, 1984.
- [67] K. A. Sharp, "Polyelectrolyte electrostatics: Salt dependence, entropic, and enthalpic contributions to free energy in the nonlinear poisson-boltzmann model," *Biopolymers*, vol. 36, pp. 227–243, 1995.
- [68] M. M. Mysinger and B. K. Shoichet, "Rapid context-dependent ligand desolvation in molecular docking," *J. Chem. Inf. Model.*, vol. 50, no. 9, pp. 1561–1573, 2010.
- [69] M. McGann, "FRED pose prediction and virtual screening accuracy," *J. Chem. Inf. Model.*, vol. 51, pp. 578–596, 2011.
- [70] H. Frauenfelder, S. Sligar, and P. Wolynes, "The energy landscapes and motions of proteins," *Science*, vol. 254, no. 5038, pp. 1598–1603, 1991.

-
- [71] P. G. Wolynes, "Recent successes of the energy landscape theory of protein folding and function," *Quarterly Reviews of Biophysics*, vol. 38, no. 4, p. 405–410, 2005.
 - [72] T. Schlick, *Molecular Modeling and Simulation*. Springer, 2010.
 - [73] G. Raab, *Molecular Simulation Studies on Thermophysical Properties With Application to Working Fluids*. Springer, 2017.
 - [74] W. Cornell, P. Cieplak, C. Bayly, I. Gould, K. Merz, Jr., D. Ferguson, D. Spellmeyer, T. Fox, J. Caldwell, and P. Kollman, "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules," *lacs*, vol. 117, pp. 5179–5197, 1995.
 - [75] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, "ff14sb: Improving the accuracy of protein side chain and backbone parameters from ff99sb," *J. Chem. Theory Comput.*, vol. 11, no. 8, pp. 3696–3713, 2015.
 - [76] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," *J. Comput. Chem.*, vol. 25, pp. 1157–1174, 2004.
 - [77] R. Salomon-Ferrer, D. a. Case, and R. C. Walker, "An overview of the amber biomolecular simulation package," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 00, no. February, pp. n/a–n/a, 2012.
 - [78] B. R. Brooks, C. L. Brooks, III, A. D. Mackerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus, "CHARMM: The biomolecular simulation program," *J. Comput. Chem.*, vol. 0, no. 0, Sp. Iss. SI, pp. 1545–1614, 2009.
 - [79] A. Cauchy, "Methodes generales pour la resolution des syst'emes dequations simultanees," *Comptes Rendus Hebd. Seances Acad. Sci*, vol. 25, pp. 536–538, 1847.
 - [80] J. C. Meza, "Steepest descent," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 6, pp. 719–722, 2010.
 - [81] R. Fletcher and C. M. Reeves, "Function minimization by conjugate gradients," *The Computer Journal*, vol. 7, no. 2, pp. 149–538, 1964.
 - [82] S. Wlodek, A. G. Skillman, and A. Nicholls, "Ligand entropy in gas-phase, upon solvation and protein complexation. fast estimation with quasi-newton hessian," *J. Chem. Theory Comput.*, vol. 6, no. 7, pp. 2140–2152, 2010.
 - [83] N. J and S. J. Wright, eds., *Numerical Optimization*. New York: Springer, 1999.
 - [84] F. E. Curtis and W. Guo, "R-linear convergence of limited memory steepest descent," *IMA Journal of Numerical Analysis*, vol. 38, no. 2, pp. 720–742, 2018.
 - [85] W. Gao and D. Goldfarb, "Quasi-newton methods: superlinear convergence without line searches for self-concordant functions," *Optimization Methods and Software*, vol. 34, no. 1, pp. 194–217, 2019.
 - [86] Q. V. Le, J. Ngiam, A. Coates, A. Lahiri, B. Prochnow, and A. Y. Ng, "On optimization methods for deep learning," *Proceedings of the 28th International Conference on Machine Learning*, pp. 265–272, 2019.
 - [87] C. Acharya, A. Coop, J. E. Polli, and D. MacKerell, "Recent advances in ligand-based drug design: Relevance and utility of the conformationally sampled pharmacophore approach," *Current Computer-Aided Drug Design*, vol. 7, no. 1, pp. 10–22, 2011.

-
- [88] Y. C. Martin, J. L. Kofron, and L. M. Traphagen, "Do structurally similar molecules have similar biological activity?," *J. Med. Chem.*, vol. 45, pp. 4350–4358, 2002.
- [89] C. Hansch, P. P. Maloney, T. Fujita, and R. M. Muir, "Correlation of biological activity of phenoxyacetic acids with hammett substituent constants and partition coefficients," *Nature*, vol. 194, pp. 178–180, 1962.
- [90] R. P. Verma and C. Hansch, "Camptothecins: A sar/qsar study," *Chem. Rev.*, vol. 109, no. 1, pp. 213–235, 2009.
- [91] G. Wolber and T. Langer, "Ligandscout: 3-d pharmacophores derived from protein-bound ligands and their use as virtual screening filters," *J. Chem. Inf. Model.*, vol. 45, no. 1, pp. 160–169, 2005.
- [92] D. Bernard, A. Coop, and A. D. MacKerell, "Conformationally sampled pharmacophore for peptidic delta opioid ligands," *J. Med. Chem.*, vol. 48, no. 24, pp. 7773–7780, 2005.
- [93] N. Nikolova and J. Jaworska, "Approaches to measure chemical similarity - a review," *QSAR. Comb. Sci.*, vol. 22, no. 9, pp. 1006–1026, 2003.
- [94] D. Rogers and M. Hahn, "Extended-connectivity fingerprints," *J. Chem. Inf. Model.*, vol. 50, no. 5, pp. 742–754, 2010.
- [95] H. L. Morgan, "The generation of a unique machine description for chemical structures-a technique developed at chemical abstracts service.," *J. Chem. Doc.*, vol. 5, no. 2, pp. 107–113, 1965.
- [96] N. M. O'Boyle, M. Banck, J. C. A. C. Morley, T. Vandermeersch, and G. R. Hutchison, "Open babel: An open chemical toolbox," *J. Cheminform.*, vol. 3, no. 33, pp. 1–14, 2011.
- [97] D. Hoover, M. Friedmann, R. Reeves, and N. S. Magnuson, "Recombinant human pim- 1 protein exhibits serine/threonine kinase activity," *J. Biol. Chem.*, vol. 25, no. 21, pp. 14018–14023, 1991.
- [98] Y. N. Zhukova, M. G. Alekseeva, N. V. Zakharevich, A. A. Shtil, and V. N. Danilenko, "Pim family of protein kinases: Structure, functions, and roles in hematopoietic malignancies," *Mol. Biol.*, vol. 45, no. 5, pp. 695–703, 2011.
- [99] J. D. Allen, E. Verhoeven, J. Domen, M. van der Valk, and A. Berns, "Pim-2 transgene induces lymphoid tumors, exhibiting potent synergy with c-myc," *Oncogene*, vol. 15, pp. 1133–1141, 1997.
- [100] N. M. T. van der Lugt, J. Domen, E. Verhoeven, K. Linders, H. van der Gulden, J. Allen, and A. Berns, "Proviral tagging in eg-myc transgenic mice lacking the pim-1 proto-oncogene leads to compensatory activation of pim-2," *EMBO J.*, vol. 14, no. 11, pp. 2536–2544, 1995.
- [101] T. Schmidt, H. Karsunky, B. Rödel, Z. B. H.-P. Elsässer, and T. Möröy, "Evidence implicating gfi-1 and pim-1 in pre-t-cell differentiation steps associated with b-selection," *EMBO J.*, vol. 17, no. 18, pp. 5349–5359, 1997.
- [102] C. Blanco-Aparicio and A. Carnero, "Pim kinases in cancer: Diagnostic, prognostic and treatment opportunities," *Biochemical Pharmacology*, vol. 85, no. 5, pp. 629–643, 2013.
- [103] J. M. Gozgit, G. Bebernitz, P. Pankaj, M. Ye, J. Parmentier, J. Wu, N. Su, T. Wang, S. Ioannidis, A. Davies, D. Huszar, and M. Zinda, "Effects of the jak2 inhibitor, az960, on pim/bad/bcl-xl survival signaling in the human jak2 v617f cell line set-2," *J. Biol. Chem.*, vol. 283, pp. 32334–32343, 2008.
- [104] N. Bhattacharya, Z. Wang, C. Davitt, I. F. McKenzie, P.-x. Xing, and N. S. Magnuson, "Pim-1 associates with protein complexes necessary for mitosis," *Chromosoma*, vol. 111, no. 2, pp. 80–95, 2002.
- [105] Z. Wang, N. Bhattacharya, P. F. Mixter, W. Wei, J. Sedivy, and N. S. Magnuson, "Phosphorylation of the cell cycle inhibitor p21cip1/waf1 by pim-1 kinase," *Biochim. Biophys. Acta*, vol. 1593, no. 1, pp. 45–55, 2002.

-
- [106] D. Morishita, R. Katayama, K. Sekimizu, T. Tsuruo, and N. Fujita, "Pim kinases promote cell cycle progression by phosphorylating and down-regulating p27kip1 at the transcriptional and posttranscriptional levels," *Cancer Research*, vol. 68, no. 13, pp. 5076–5085, 2008.
 - [107] T. Mochizuki, C. Kitanaka, K. Noguchi, T. Muramatsu, A. Asai, and Y. Kuchino, "Physical and functional interactions between pim-1 kinase and cdc25a phosphatase," *J. Biol. Chem.*, vol. 274, no. 26, pp. 18659–18666, 1999.
 - [108] M. Bachmann, C. Kosan, X. X. P, M. Montenarh, I. Hoffmann, and T. Möröy, "The oncogenic serine/threonine kinase pim-1 directly phosphorylates and activates the g2/m specific phosphatase cdc25c," *Int. J. Biochem. Cell. Biol.*, vol. 38, no. 3, pp. 430–443, 2006.
 - [109] K.-T. Kim, M. Levis, and D. Small, "Constitutively activated flt3 phosphorylates bad partially through pim-1," *British Journal of Haematology*, vol. 134, pp. 500–209, 2006.
 - [110] V. H. Cowling and M. H. Cole, "Mechanism of transcriptional activation by the myc oncoproteins," *Seminars in Cancer Biology*, vol. 16, pp. 242–252, 2006.
 - [111] M. Vita and M. Henriksson, "The myc oncoprotein as a therapeutic target for human cancer," *Seminars in Cancer Biology*, vol. 16, no. 4, pp. 318–330, 2006.
 - [112] E. Yeh, M. Cunningham, H. Arnold, D. Chasse, T. Monteith, G. Ivaldi, W. C. Hahn, P. T. Stukenberg, S. Shenolikar, T. Uchida, C. M. Counter, J. R. Nevins, A. R. Means, and R. Sears, "A signalling pathway controlling c-myc degradation that impacts oncogenic transformation of human cells," *Nat. Cel. Biol*, vol. 6, pp. 308–318, 2004.
 - [113] S. Verbeek, M. van Lohuizen, M. van der Valk, J. Domen, G. Kraal, and A. Berns, "Mice bearing the e mu-myc and e mu-pim-1 transgenes develop pre-b-cell leukemia prenatally.," *Mol. Cell. Biol.*, vol. 11, no. 2, pp. 1176–1179, 1991.
 - [114] Y. Zhang, Z. Wang, X. Li, and N. S. Magnuson, "Pim kinase-dependent inhibition of c-myc degradation," *Oncogene*, vol. 27, pp. 4809–4819, 2008.
 - [115] J. Wang, J. Kim, M. Roh, O. E. Franco, S. W. Hayward, M. L. Wils, and S. A. Abdulkadir, "Pim1 kinase synergizes with c-myc to induce advanced prostate carcinoma," *Oncogene*, vol. 29, pp. 2477–2487, 2010.
 - [116] M. C. Nawijn, A. J. Alendar, and A. Berns, "For better or for worse: the role of pim oncogenes in tumorigenesis," *Nat. Rev. Cancer*, vol. 11, pp. 23–24, 2011.
 - [117] P. D. Garcia, J. L. Langowski, Y. Wang, M. Chen, J. Castillo, C. Fanton, M. Ison, T. Zavorotinskaya, Y. Dai, J. Lu, X.-H. Niu, S. Basham, J. Chan, J. Yu, M. Doyle, P. Feucht, R. Warne, J. Narberes, T. Tsang, C. Fritsch, A. Kauffmann, E. Pfister, P. Drueckes, J. Trappe, C. Wilson, W. Han, J. Lan, G. Nishiguchi, M. Lindvall, C. Bellamacina, J. A. Aycinena, R. Zang, J. Holash, and M. T. Burger, "Pan-pim kinase inhibition provides a novel therapy for treating hematologic cancers," *Clin. Canc. Res.*, vol. 20, no. 7, pp. 1834–1845, 2014.
 - [118] A. C. Carrera, K. Alexandrov, and T. M. Roberts, "The conserved lysine of the catalytic domain of protein kinases is actively involved in the phosphotransfer reaction and not required for anchoring atp," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 90, no. 2, pp. 442–446, 1993.
 - [119] M. Cherry and D. H. Williams, "Recent kinase and kinase inhibitor x-ray structures: Mechanisms of inhibition and selectivity insights," *Curr. Med. Chem.*, vol. 11, pp. 663–673, 2004.
 - [120] F. Anizon, A. A. Shtil, V. N. Danilenko, and P. Moreau, "Fighting tumor cell survival: Advances in the design and evaluation of pim inhibitors," *Curr. Med. Chem.*, vol. 17, pp. 4114–4133, 2010.

- [121] Z. A. Knight and K. M. Shokat, "Features of selective kinase inhibitors," *Chemistry and Biology*, vol. 12, pp. 621–637, 2005.
- [122] M. D. Jacobs, J. Black, O. Futer, L. Swenson, B. Hare, M. Fleming, and K. Saxena, "Pim-1 ligand-bound structures reveal the mechanism of serine/threonine kinase inhibition by ly294002," *J. Biol. Chem.*, vol. 280, no. 14, pp. 13728–13734, 2005.
- [123] M. T. Burger, W. Han, J. Lan, G. Nishiguchi, C. Bellamacina, M. Lindval, G. Atallah, Y. Ding, M. Mathur, C. McBride, E. L. Beans, K. Muller, V. Tamez, Y. Zhang, K. Huh, P. Feucht, T. Zavorotinskaya, Y. Dai, J. Holash, J. Castillo, J. Langowski, Y. Wang, M. Y. Chen, and P. D. Garcia, "Structure guided optimization, in vitro activity, and in vivo activity of pan-pim kinase inhibitors," *ACS Med. Chem. Lett.*, vol. 4, no. 12, pp. 1193–1197, 2013.
- [124] A. Ishchenko, L. Zhang, J.-Y. Le Brazidec, J. Fan, J. H. Chong, A. Hingway, A. Raditsis, L. Singh, B. Elenbaas, V. S. Hong, *et al.*, "Structure-based design of low-nanomolar pim kinase inhibitors," *Bioorg. Med. Chem. Lett.*, vol. 25, no. 3, pp. 474–480, 2015.
- [125] R. Bohacek, C. McMartin, and W. Guida, "The art and practice of structure-based drug design: A molecular modeling perspective," *Med. Res. Rev.*, vol. 16, no. 1, pp. 3–50, 1996.
- [126] R. J. Hall, P. N. Mortenson, and C. W. Murray, "Efficient exploration of chemical space by fragment-based screening," *Progr. Biophys. Mol. Biol.*, vol. 116, no. 2, pp. 82–91, 2014.
- [127] J.-L. Reymond, "The chemical space project," *Acc. Chem. Res.*, vol. 48, pp. 722–730, 2015.
- [128] I. Akritopoulou-Zanze and P. J. Hajduk, "Kinase-targeted libraries: The design and synthesis of novel, potent, and selective kinase inhibitors," *Drug Discov. Today*, vol. 14, no. 5, pp. 291–297, 2009.
- [129] A. S. Karpov, P. Amiri, C. Bellamacina, M.-H. Bellance, W. Breitenstein, D. Daniel, R. Denay, D. Fabbro, C. Fernandez, I. Galuba, S. Guerre-Lagasse, S. Gutmann, L. Hinh, W. Jahnke, J. Klopp, A. Lai, M. K. Lindvall, S. Ma, H. Mbitz, S. Pecchi, G. Rummel, K. Shoemaker, J. Trappe, C. Voliva, S. W. Cowan-Jacob, and A. L. Marzinzik, "Optimization of a dibenzodiazepine hit to a potent and selective allosteric pak1 inhibitor," *ACS Med. Chem. Lett.*, vol. 6, no. 7, pp. 776–781, 2015.
- [130] G. M. Keser, D. A. Erlanson, G. G. Ferenczy, M. M. Hann, C. W. Murray, and S. D. Pickett, "Design principles for fragment libraries—maximizing the value of learnings from pharma fragment based drug discovery (fbdd) programs for use in academia," *J. Med. Chem.*, 2016.
- [131] J. J. Irwin, T. Sterling, M. M. Mysinger, E. S. Bolstad, and R. G. Coleman, "ZINC: A free tool to discover chemistry for biology," *J. Chem. Inf. Model.*, vol. 52, no. 7, pp. 1757–1768, 2012.
- [132] A. Widmer, "WITNOTP : A computer program for molecular modelling," 1997.
- [133] I. D. Kuntz, E. C. Meng, S. J. Oatley, R. Langridge, and T. E. Ferrin, "A geometric approach to macromolecule-ligand interactions," *J. Mol. Biol.*, vol. 161, pp. 269–288, 1982.
- [134] D. Weininger, "Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules," *J. Chem. Inf. Comput. Sci.*, vol. 28, no. 1, pp. 31–36, 1988.
- [135] D. Weininger, A. Weininger, and J. L. Weininger, "Smiles. 2. algorithm for generation of unique smiles notation," *J. Chem. Inf. Comput. Sci.*, vol. 29, no. 2, pp. 97–101, 1989.
- [136] D. Weininger, "Smiles. 3. depict. graphical depiction of chemical structures," *J. Chem. Inf. Comput. Sci.*, vol. 30, no. 3, pp. 237–243, 1990.
- [137] N. OpenEye Scientific Software, Santa Fe, "QUACPAC 1.7.0.2." (accessed Nov 15, 2014).

-
- [138] P. C. D. Hawkins, A. G. Skillman, G. L. Warren, B. A. Ellingson, and M. T. Stahl, "Conformer generation with OMEGA: Algorithm and validation using high quality structures from the protein databank and cambridge structural database," *J. Chem. Inf. Model.*, vol. 50, no. 4, pp. 572–584, 2010.
- [139] A. Jakalian, D. B. Jack, and C. I. Bayly, "Fast, efficient generation of high-quality atomic charges. am1-bcc model: Ii. parameterization and validation," *J. Comput. Chem.*, vol. 23, pp. 1623–1641, 2002.
- [140] C. Axon, "Calculator Plugins 16.1.11." (accessed Jan 11, 2016).
- [141] J. Sadowski, J. Gasteiger, and K. G., "Comparison of automatic three-dimensional model builders using 639 x-ray structures," *J. Chem. Inf. Comput. Sci.*, vol. 34, pp. 1000–1008, 1994.
- [142] G. Hawkins, D. Giesen, G. Lynch, C. Chambers, and I. e. a. Rossi, "Amsol-version 7.1."
- [143] N. Miyaoura, K. Yamada, and A. Suzuki, "A new stereospecific cross-coupling by the palladium-catalyzed reaction of 1-alkenylboranes with 1-alkenyl or 1-alkynyl halides," *Tetrahedron Letters*, vol. 36, pp. 3437–3440, 1979.
- [144] Molport, "Molport building blocks." (accessed Sep 24, 2018).
- [145] D. A. Case, V. Bablin, J. T. Berryman, *et al.*, "Amber 14, university of california, san francisco," 2014.
- [146] D. A. Case, R. M. Betz, D. S. Cerutti, *et al.*, "Amber 16, university of california, san francisco," 2016.
- [147] C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman, "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The resp model," *J. Phys. Chem.*, vol. 97, pp. 10269–10280, 1993.
- [148] M. J. Frisch, G. W. Trucks, H. B. Schlegel, *et al.*, "Gaussian 09, revision c.01," 2010.
- [149] J. Wang, W. Wang, K. P. A, and D. A. Case, "Automatic atom type and bond type perception in molecular mechanical calculations," *J. Mol. Graphics Modell.*, vol. 25, pp. 247–260, 2006.
- [150] C. H. R. Chemical Computing Group ULC 1010 Sherbooke St. West Suite 910, Montreal QC, "Molecular operating environment (moe)," 2018.
- [151] V. Hornak, R. Abel, A. Okur, B. Strockbine, R. A, and C. Simmerling, "Comparison of multiple amber force fields and development of improved protein backbone parameters," *Proteins: Struct., Funct., Bioinf.*, vol. 25, pp. 712–725, 2006.
- [152] W. L. Jorgenson, J. Chandrasekhar, J. D. Madura, R. W. Impey, and K. M. I, "Comparison of simple potential functions for simulating liquid water," *J. Chem. Phys.*, vol. 79, no. 2, pp. 926–935, 1983.
- [153] J.-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes," *J. Chem. Phys.*, vol. 23, pp. 327–341, 1977.
- [154] R. W. Pastor, B. R. Brooks, and A. Szabo, "An analysis of the accuracy of langevin and molecular dynamics algorithms," *Mol. Phys.*, vol. 65, no. 6, pp. 1409–1419, 1983.
- [155] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, "Improved side-chain torsion potentials for the amber ff99sb protein force field," *Proteins: Struct., Funct., Bioinf.*, vol. 78, pp. 1950–1958, 2010.
- [156] N. OpenEye Scientific Software, Santa Fe, "SZYBKI 1.8.0.2." (accessed Nov 15, 2014).
- [157] T. A. Halgren, "Merck molecular force field. i. basis, form, scope, parameterization, and performance of mmff94," *J. Comput. Chem.*, vol. 17, no. 5, pp. 490–519, 1996.

-
- [158] B. Honig, K. Sharp, and A.-S. Yang, "Macroscopic models of aqueous solutions: Biological and chemical applications," *J. Phys. Chem.*, vol. 97, pp. 1101–1109, 1993.
- [159] D. C. Liu and J. Nocedal, "On the limited memory bfgs method for large scale optimization," *Mathematical Programming*, vol. 45, pp. 503–528, 1989.
- [160] T. W. H. Backman, Y. Cao, and T. Girke, "Chemmine tools: an online service for analyzing and clustering small molecules," *Nucleic Acids Res.*, vol. 39, pp. W486–W491, 2011.
- [161] "Qtplot, a program for two- and three-dimensional graphical presentation of data sets and for data analysis."
- [162] F. J. Richards, "A flexible growth function for empirical use," *Journal of Experimental Botany*, vol. 10, no. 29, pp. 290–300, 1959.
- [163] C. Yung-Chi and W. H. Prusoff, "Relationship between the inhibition constant (k_i) and the concentration of inhibitor which causes 50 per cent inhibition (i_{50}) of an enzymatic reaction," *Biochemical Pharmacology*, vol. 22, no. 23, pp. 3099 – 3108, 1973.
- [164] J. J. Irwin, D. Duan, H. Torosyan, A. K. Doak, K. T. Ziebart, T. Sterling, G. Tumanian, and B. K. Shoichet, "An aggregation advisor for ligand discovery," *J. Med. Chem.*, vol. 58, pp. 7076–7087, 2015.
- [165] C. Yung-Chi and W. H. Prusoff, "Relationship between the inhibition constant (k_i) and the concentration of inhibitor which causes 50 per cent inhibition (i_{50}) of an enzymatic reaction," *Biochem. Pharm.*, vol. 22, no. 23, pp. 3099–3108, 1973.
- [166] K. Babaoglu, A. Simeonov, J. J. Lrwin, M. E. Nelson, B. Feng, C. J. Thomas, L. Cancian, M. P. Costi, D. A. Maltby, A. Jadhav, J. Inglese, C. P. Austin, and B. K. Shoichet, "Comprehensive mechanistic analysis of hits from high-throughput and docking screens against beta-lactamase," *J. Med. Chem.*, vol. 51, no. 8, pp. 2502–2511, 2008.
- [167] M. M. Mysinger, D. R. Weiss, J. J. Ziarek, S. Gravel, A. K. Doak, J. Karpiak, N. Heveker, B. K. Shoichet, and B. F. Volkman, "Structure-based ligand discovery for the protein–protein interface of chemokine receptor cxcr4," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 14, pp. 5517–5522, 2012.
- [168] I. Müller, "Guidelines for the successful generation of protein–ligand complex crystals," *Acta Cryst. D. Struct. Biol.*, vol. 73, no. 2, pp. 79–92, 2016.
- [169] J. E. Hutti, E. T. Jarrell, J. D. Chang, D. W. Abbott, P. Storz, T. A. L. C. Cantley, and T. B. E, "A rapid method for determining protein kinase phosphorylation specificity," *Nature Methods*, vol. 1, pp. 27–29, 2004.
- [170] Z. Ul-Haq, S. Gul, S. Usmani, A. Wadood, and W. Khan, "Binding site identification and role of permanent water molecule of pim-3 kinase: A molecular dynamics study," *J. Mol. Graphics Modell.*, vol. 62, pp. 276–282, 2015.
- [171] J. B. Cross, D. C. Thompson, B. K. Rai, J. C. Babr, K. Y. Fan, Y. Hu, and C. Humblet, "Comparison of several molecular docking programs: Pose prediction and virtual screening accuracy," *J. Chem. Inf. Model.*, vol. 49, pp. 1455–1474, 2009.
- [172] K. T. Debiec, A. M. Gronenborn, and L. T. Chong, "Evaluating the strength of salt bridges: A comparison of current biomolecular force fields," *J. Phys. Chem. B*, vol. 118, no. 24, pp. 6561–6569, 2014.
- [173] A. Rubenstein and S. Sherman, "Influence of the solvent structure on the electrostatic interactions in proteins," *Biophys. J.*, vol. 87, pp. 1544–1557, 2004.
- [174] M. K. Gilson, A. Rashin, R. Fine, and H. B, "On the calculation of electrostatic interactions in proteins," *J. Mol. Biol.*, vol. 183, pp. 503–516, 1985.

-
- [175] H. Eldar-Finkelman and M. Eisenstein, "Peptide inhibitors targeting protein kinases," *Current Pharmaceutical Design*, vol. 15, pp. 2463–2470, 2009.
- [176] Q. Shen, V. S. Bhatt, I. Krieger, J. C. Sacchettini, and J.-H. Cho, "Structure-guided design of a potent peptide inhibitor targeting the interaction between crk and abl kinase," *Med. Chem. Comm.*, vol. 9, pp. 519–524, 2018.
- [177] H.-L. Wang, K. L. Andrews, S. K. Booker, J. Canon, V. J. Cee, F. Chavez, Y. Chen, H. Eastwood, N. Guerrero, B. Herberich, D. Hickman, B. A. Lanman, J. Laszlo, M. R. Lee, J. R. Lipford, B. Mattson, C. Mohr, Y. Nguyen, M. H. Norman, L. H. Pettus, D. Powers, A. B. Reed, K. Rex, C. Sastri, N. Tamayo, P. Wang, J. T. Winston, B. Wu, Q. Wu, T. Wu, R. P. Wurcz, Y. Xu, Y. Zhou, and A. S. Tasker, "Discovery of (r)-8-(6-methyl-4-oxo-1,4,5,6-tetrahydropyrrolo[3,4-b]pyrrol-2-yl)-3-(1-methylcyclopropyl)-2-((1-methylcyclopropyl)amino)quinazolin-4(3h)-one, a potent and selective pim-1/2 kinase inhibitor for hematological malignancies," *J. Med. Chem.*, vol. 62, no. 3, pp. 1523–1540, 2019.
- [178] C. Watanabe, H. Watanabe, K. Fukuzawa, L. J. Parker, Y. Okiyama, H. Yuki, S. Yokoyama, H. Nakano, S. Tanaka, and T. Honma, "Theoretical analysis of activity cliffs among benzofuranone-class pim1 inhibitors using the fragment molecular orbital method with molecular mechanics poisson-boltzmann surface area (fmo+mm-pbsa) approach," *J. Chem. Inf. Model.*, vol. 57, no. 12, pp. 2996–3010, 2017.
- [179] X. Wang, W. Blackaby, V. Allen, G. K. Y. Chan, J. H. Chang, P.-C. Chiang, C. Diane, J. Drummond, S. Do, E. Fan, E. B. Harstad, A. Hodges, H. Hu, W. Jia, W. Kofie, A. Kolesnikov, J. P. Lyssikatos, J. Ly, M. Matteucci, J. G. Moffat, V. Munugalavadla, J. Murray, D. Nash, C. L. Noland, G. Del Rosario, L. Ross, C. Rouse, A. Sharpe, D. Slaga, M. Sun, V. Tsui, H. Wallweber, S.-F. Yu, and A. J. Ebens, "Optimization of pan-pim kinase activity and oral bioavailability leading to diaminopyrazole (gdc-0339) for the treatment of multiple myeloma," *J. Med. Chem.*, vol. 62, no. 4, pp. 2140–2153, 2019.
- [180] P. W. Ingham and A. P. McMahon, "Hedgehog signaling in animal development: paradigms and principles," *Genes and Development*, vol. 15, pp. 3057–3087, 2001.
- [181] H. Zhang, Z. Sun, Z. Liu, and C. Song, "Overcoming the emerging drug resistance of smoothened: an overview of small-molecule smo antagonists with antiresistance activity," *Future Medicinal Chemistry*, vol. 10, no. 24, pp. 2855–2875, 2018.
- [182] J. Briscoe and P. P. Thérond, "The mechanisms of hedgehog signalling and its roles in development and disease," *Nat. Rev. Mol. Cell Biol.*, vol. 14, pp. 416–429, 2013.
- [183] S. Teglund and R. Toftgård, "Hedgehog beyond medulloblastoma and basal cell carcinoma," *Biochim. Biophys. Acta*, vol. 1805, no. 2, pp. 181–208, 2010.
- [184] A. M. Arensdorf, S. Marada, and S. K. Ogden, "Smoothened regulation: A tale of two signals," *Trends Pharmacol. Sci.*, vol. 37, no. 1, pp. 62–72, 2016.
- [185] S. M. Foord, T. I. Bonner, R. R. Neubig, E. M. Rosser, J.-P. Pin, A. P. Davenport, M. Spedding, and A. J. Harmar, "International union of pharmacology. xlvii. g protein-coupled receptor list," *Pharmacol. Rev.*, vol. 57, no. 2, pp. 279–288, 2005.
- [186] M. Philipp and M. G. Caron, "Hedgehog signalling: Is smo a g protein-coupled receptor?," *Curr Biol.*, vol. 19, no. 3, pp. R125–R127, 2008.
- [187] S. K. Ogden, D. L. Fei, N. S. Schilling, Y. F. Ahmed, J. Hwa, and D. J. Robbins, "G protein gai functions immediately downstream of smoothened in hedgehog signalling," *Nature*, vol. 456, pp. 967–970, 2008.
- [188] M. He, R. Subramanian, F. Bangs, F. Omelchenko, K. F. Liem, K. T. M., and K. V. Anderson, "The kinesin-4 protein kif7 regulates mammalian hedgehog signalling by organizing the cilium tip compartment," *Nat. Cel. Biol.*, vol. 16, pp. 663–672, 2014.

-
- [189] T. Gorojankina, "Hedgehog signaling pathway: a novel model and molecular mechanisms of signal transduction," *Cell. Mol. Life. Sci.*, vol. 73, no. 7, pp. 1317–1332, 2016.
- [190] S. Gupta, N. Takebe, and P. Lorusso, "Targeting the hedgehog pathway in cancer," *Ther. Adv. Med. Oncol*, vol. 2, no. 4, pp. 237–250, 2010.
- [191] H. J. Sharpe, G. Pau, G. J. Dijkgraaf, N. Basset-Seguin, Z. Modrusan, T. Januario, V. Tsui, A. B. Durham, A. A. Dlugosz, P. M. Haverty, R. Bourgon, J. Y. Tang, K. Y. Sarin, L. Dirix, D. C. Fisher, C. M. Rudin, H. Sofen, M. R. Migden, R. L. Yauch, and F. A. de Sauvage, "Genomic analysis of smoothened inhibitor resistance in basal cell carcinoma," *Cancer Cell*, vol. 27, no. 3, pp. 327–341, 2015.
- [192] G. J. P. Dijkgraaf, B. Aliche, L. Weinmann, T. Januario, K. West, Z. Modrusan, D. Burdick, R. Goldsmith, K. Robarge, D. Sutherlin, S. J. Scales, S. E. Gould, R. L. Yauch, and F. J. de Sauvage, "Small molecule inhibition of gdc-0449 refractory smoothened mutants and downstream mechanisms of drug resistance," *Cancer Research*, vol. 71, no. 2, pp. 435–444, 2011.
- [193] A. S. Kiselyov, "Targeting the hedgehog signaling pathway with small molecules," *Ant. Canc. Agents In Med. Chem.*, vol. 6, pp. 445–449, 2006.
- [194] C. Wang, W. Huixian, V. Katritch, G. W. Han, X.-P. Huang, W. Liu, F. Y. Siu, B. L. Roth, V. Cherezov, and R. Stephens, "Structure of the human smoothened receptor bound to an antitumour agent," *Nature*, vol. 497, pp. 338–343, 2013.
- [195] D. D. Von Hoff, P. M. LoRusso, C. M. Rudin, J. C. Reddy, R. L. Yauch, R. Tibes, G. J. Weiss, M. J. Borad, C. L. Hann, J. R. Brahmer, H. M. Mackey, B. L. Lum, W. C. Darbonne, J. C. Marsters, F. J. de Sauvage, and J. A. Low, "Inhibition of the hedgehog pathway in advanced basal-cell carcinoma," *New England J. Med.*, vol. 361, no. 12, pp. 1164–1172, 2009.
- [196] R. L. Yauch, G. J. P. Dijkgraaf, B. Aliche, T. Januario, C. P. Ahn, T. Holcomb, K. Pujara, J. Stinson, C. A. Callahan, T. Tang, J. F. Bazan, Z. Kan, S. Seshagiri, C. L. Hann, S. E. Gould, J. A. Low, C. M. Rudin, and F. J. de Sauvage, "Smoothened mutation confers resistance to a hedgehog pathway inhibitor in medulloblastoma," *Science*, vol. 326, no. 5952, pp. 572–574, 2009.
- [197] M. H. Bender, P. A. Hipskind, A. R. Capen, M. Cockman, K. M. Credille, H. Gao, J. A. Bastian, J. M. Clay, K. L. Lobb, D. J. Sall, M. L. Thompson, T. Wilson, G. N. Wishart, and B. K. Patel, "Abstract 2819: Identification and characterization of a novel smoothened antagonist for the treatment of cancer with deregulated hedgehog signaling," *Cancer Research*, vol. 71, no. 8 Supplement, pp. 2819–2819, 2011.
- [198] C. Wang, W. Huixian, T. Evron, G. W. Han, X.-P. Huang, S. J. Hufeisen, T. J. Mangano, D. J. Urban, V. Katritch, , V. Cherezov, C. M. G, B. L. Roth, and R. Stephens, "Structural basis for smoothened receptor modulation and chemoresistance to anticancer drugs," *Nature Communications*, vol. 5, no. 4355, pp. 1–11, 2014.
- [199] S. Atwood, K. A. Y. Sarin, R. J. Whitson, J. R. Li, G. Kim, M. Rezaee, M. S. Ally, J. Kim, C. Yao, A. L. S. Chang, A. E. Oro, and J. Y. Tang, "Smoothened variants explain the majority of drug resistance in basal cell carcinoma," *Cancer Cell*, vol. 27, no. 3, pp. 342–353, 2015.
- [200] X. Zhang, F. Zhao, Y. Wu, J. Wang, W. Han, S. Zhao, A. Ishchenko, Y. Lintao, X. Lin, K. Ding, V. Dharmarajan, P. J. Griffin, C. Gato, G. Nelson, M. S. Hunter, M. A. Hanson, V. Cherezov, R. C. Stephens, W. Tan, H. Tao, and F. Xu, "Crystal structure of a multi-domain human smoothened receptor in complex with a super stabilizing ligand," *Nature Communications*, vol. 8, p. 15383, 2016.
- [201] E. Byrne, R. Sircar, P. S. Miller, G. Hedger, G. Luchetti, S. Nachtergaele, M. D. Tully, L. Mydock-McGrane, D. F. Covey, R. P. Rambo, S. M. S, N. S, R. Rohatgi, and S. C, "Structural basis of smoothened regulation by its extracellular domains," *Nature*, vol. 535, pp. 517–522, 2016.

-
- [202] B. R. Myers, L. Neahring, Y. Zhang, K. J. Roberts, and P. A. Beachy, "Rapid, direct activity assays for smoothened reveal hedgehog pathway regulation by membrane cholesterol and extracellular sodium," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 114, no. 52, pp. E11141–E11150, 2017.
- [203] G. Luchetti, R. Sircar, J. H. Kong, S. Nachtergaele, A. Sagner, E. F. Byrne, D. F. Covey, C. Siebold, and R. Rohatgi, "Cholesterol activates the g-protein coupled receptor smoothened to promote hedgehog signaling," *eLife*, vol. 5, p. e20304, 2016.
- [204] K. No, J. Grant, and H. Scheraga, "Determination of net atomic charges using a modified partial equalization of orbital electronegativity method. 1. Application to neutral molecules as models for polypeptides," *J. Phys. Chem.*, vol. 94, pp. 4732–4739, 1990.
- [205] D. A. Case, T. E. Cheatham, T. Darden, H. Göhlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, "The amber biomolecular simulation programs," *J. Comput. Chem.*, vol. 26, pp. 1668–1688, 2005.
- [206] P. R. Gerber and Müller, "Mab, a generally applicable molecular force field for structure modelling in medicinal chemistry," *J. Comput.-Aided Mol. Design*, vol. 9, no. 3, pp. 251–268, 1995.
- [207] P. Labute, "The moe nonlinear optimization library," 1995.
- [208] M. M. Mysinger, M. Carchia, J. J. Irwin, and B. K. Shoichet, "Directory of useful decoys, enhanced (dud-e): better ligands and decoys for better benchmarking," *J. Med. Chem.*, vol. 55, no. 14, pp. 6582–6594, 2012.
- [209] D. S. Wishart, T. Jewison, A. C. Guo, M. Wilson, C. Knox, Y. Liu, Y. Djoumbou, R. Mandal, F. Aziat, E. Dong, S. Bouatra, I. Sinelnikov, D. Arndt, J. Xia, P. Liu, F. Yallou, T. Bjorn Dahl, R. Perez-Pineiro, R. Eisner, F. Allen, V. Neveu, R. Greiner, and A. Scalbert, "Hmdb 3.0-the human metabolome database in 2013," *Nucleic Acids Res.*, vol. 41, pp. D801–D807, 2012.
- [210] M. Kanehisa and S. Goto, "Kegg: Kyoto encyclopedia of genes and genomes," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 27–30, 2000.
- [211] I. J. Bruno, J. C. Cole, M. Kessler, J. Luo, W. D. S. Motherwell, L. H. Purkis, B. R. Smith, and R. Taylor, "Retrieval of crystallographically-derived molecular geometry information," *J. Chem. Inf. Model.*, vol. 44, pp. 2133–2144, 2004.
- [212] C. Lacroix, I. Fish, H. Torosyan, P. Parathaman, J. J. Irwin, B. K. Shoichet, and S. Angers, "Identification of novel smoothened ligands using structure-based docking," *PLOS ONE*, vol. 11, pp. 1–20, 08 2016.
- [213] H. J. Sharpe and F. J. de Sauvage, "An oxysterol ligand for smoothened," *Nat. Chem. Biol.*, vol. 8, pp. 139–140, 2012.
- [214] E. F. X. Byrne, G. Luchetti, R. Rohatgi, and C. Siebold, "Multiple ligand binding sites regulate the hedgehog signal transducer smoothened in vertebrates," *Curr. Opin. Cell. Biol.*, vol. 51, pp. 81–88, 2018.
- [215] R. W. King, "Roughing up smoothened: chemical modulators of hedgehog signaling," *Journal of Biology*, vol. 1, no. 2, pp. 1–8, 2002.
- [216] U. Weierstall, D. James, C. Wang, T. A. White, D. Wang, W. Liu, J. C. H. Spence, R. B. Doak, G. Nelson, P. Fromme, R. Fromme, I. Grotjohann, C. Kupitz, N. A. Zatsepin, H. Liu, S. Basu, D. Wacker, G. W. Han, V. Katritch, S. Boutet, M. Messerschmidt, G. J. Williams, J. E. Koglin, M. M. Seibert, M. Klinker, C. Gati, R. L. Shoeman, A. Barty, H. N. Chapman, R. A. Kirian, K. R. Beyerlein, R. C. Stevens, D. Li1, S. T. A. Shah, N. Howe11, M. Caffrey, and V. Cherezov, "Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography," *Nature Communications*, vol. 5, no. 3309, pp. 1–6, 2013.

-
- [217] D. R. Raleigh, N. Sever, P. K. Choksi, M. A. Sigg, K. M. Hines, B. M. Thompson, D. Elnatan, P. Jaisankar, P. Bisignano, F. R. Garcia-Gonzalo, A. L. Krup, M. Eberl, E. F. Byrne, C. Siebold, S. Y. Wong, A. R. Renslo, M. Grabe, J. G. McDonald, L. Xu, P. A. Beachy, and J. F. Reiter, "Cilia-associated oxysterols activate smoothened," *Mol. Cell*, vol. 72, no. 2, pp. 316 – 327.e5, 2018.
- [218] K. K. Chahal, M. Parle, and R. Abagyan, "Hedgehog pathway and smoothened inhibitors in cancer therapies," *Anti-Cancer Drugs*, vol. 29, no. 5, pp. 387–401, 2018.
- [219] P. Kolb, D. M. Rosenbaum, J. J. Irwin, J. J. Fung, B. K. Kobilka, and B. K. Shoichet, "Structure-based discovery of β_2 -adrenergic receptor ligands," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, pp. 6843–6848, APR 21 2009.
- [220] J. Carlsson, R. G. Coleman, V. Setola, J. J. Irwin, H. Fan, A. Schlessinger, A. Sali, B. L. Roth, and B. K. Shoichet, "Ligand discovery from a dopamine D(3) receptor homology model and crystal structure," *Nat. Chem. Biol.*, vol. 7, no. 11, pp. 769–778, 2011.
- [221] S. Song, J. Jiang, L. Zhao, Q. Wang, W. Lu, C. Zheng, J. Zhang, H. Ma, S. Tian, J. Zheng, L. Luo, Y. Li, Z.-J. Yang, and X. Zhang, "Structural optimization on a virtual screening hit of smoothened receptor," *Eur. J. Med. Chem.*, vol. 172, pp. 1–15, 2019.
- [222] T. Pantsar and A. Poso, "Binding affinity via docking: Fact and fiction," *Molecules*, vol. 23, no. 8, 2018.
- [223] A. M. Ruvinsky, "Role of binding entropy in the refinement of protein–ligand docking predictions: Analysis based on the use of 11 scoring functions," *J. Comput. Chem.*, vol. 28, no. 8, pp. 1364–1372, 2007.
- [224] C.-e. a. Chang, W. Chen, and M. K. Gilson, "Ligand configurational entropy and protein binding.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 5, pp. 1534–9, 2007.
- [225] F. Manzonni and U. Ryde, "Assessing the stability of free-energy perturbation calculations by performing variations in the method," *J. Comput.-Aided Mol. Design*, vol. 32, no. 4, pp. 529–536, 2018.
- [226] J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, M. Head-Gordon, G. N. I. Clark, M. E. Johnson, and T. Head-Gordon, "Current status of the amoeba polarizable force field," *J. Phys. Chem. B*, vol. 114, no. 8, pp. 2549–2564, 2010.
- [227] D. Yusuf, A. M. Davis, G. J. Kleywegt, and S. Schmitt, "An alternative method for the evaluation of docking performance: Rsr vs rmsd," *J. Chem. Inf. Model.*, vol. 48, no. 7, pp. 1411–1422, 2008.

Chapter 7

Appendix

7.1 Novel ligands against the kinase Proviral Integration site for MuLV (Pim-1)

7.1.1 Project 1: Docking software comparisons vs structures from MD simulation

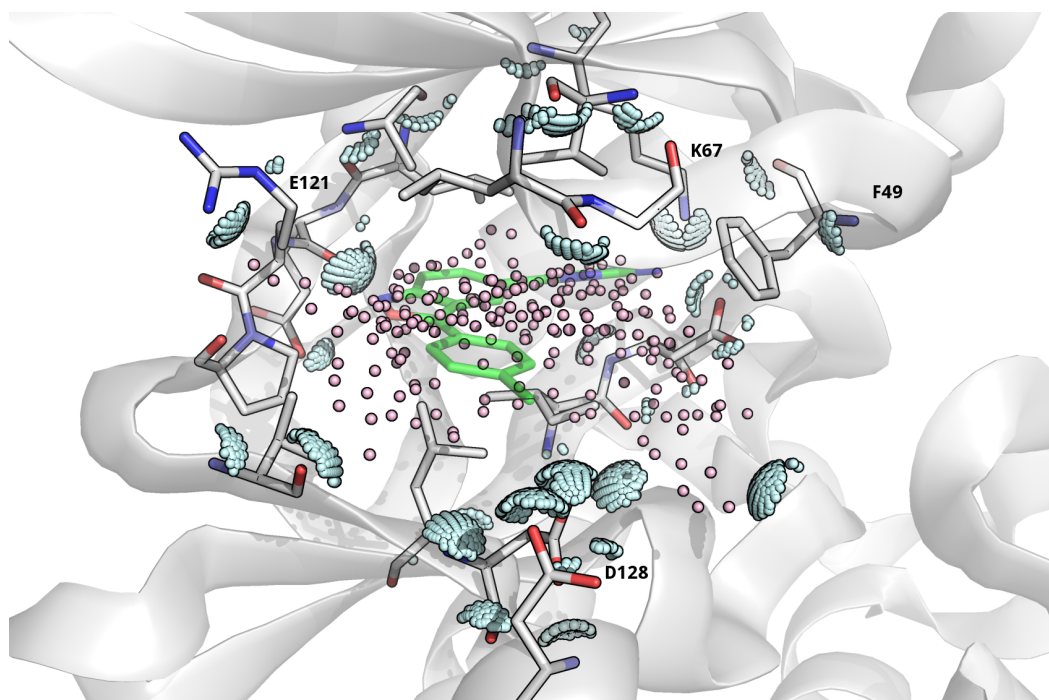


Figure 7.1: Prepared Pim-1 docking binding site polar (light blue, small spheres) and apolar (light pink, small spheres) vectors for SEED with crystal ligand (green carbons, stick representation) overlay (PDB:3BGP). Key residues (light grey, stick representation) are labelled.

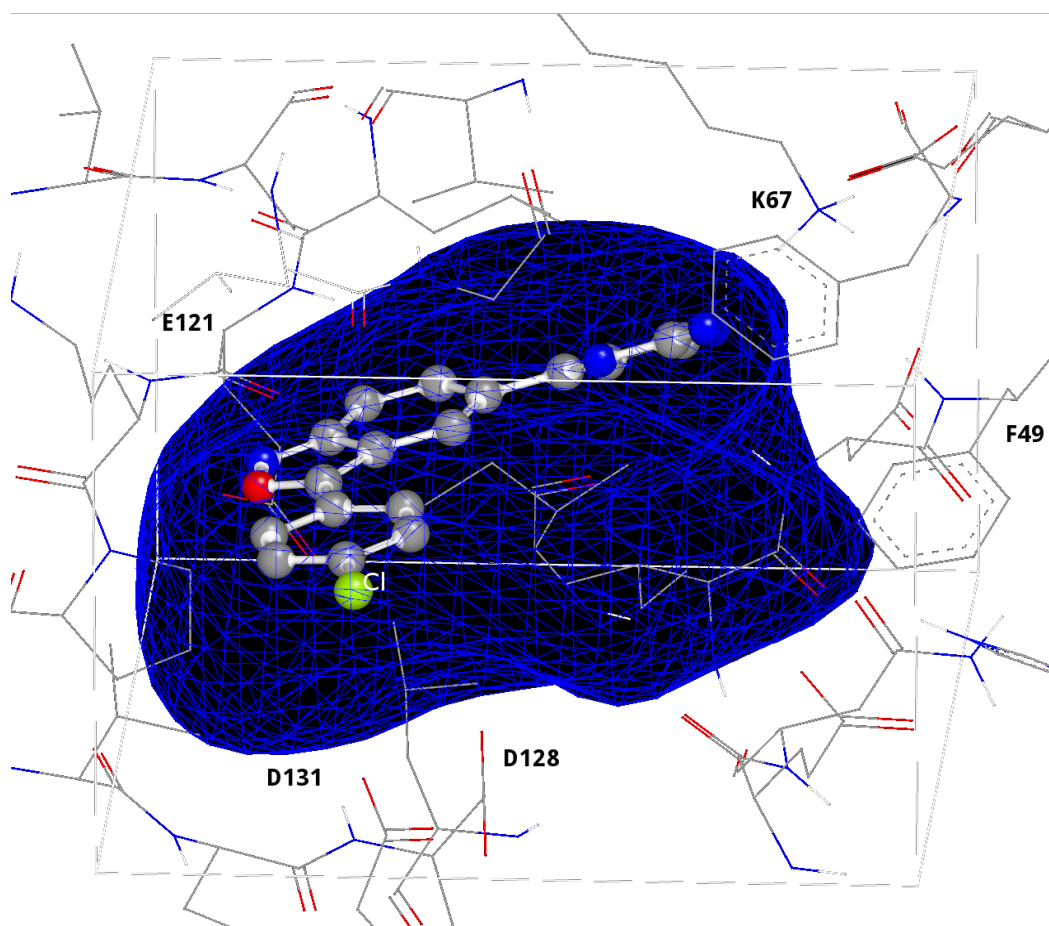


Figure 7.2: Prepared Pim-1 docking binding sites for FRED with binding site contours (dark blue, mesh representation) with crystal ligand (grey carbons, ball-and-stick representation) overlay (PDB:3BGP). Key residues (light grey, stick representation) are labelled.

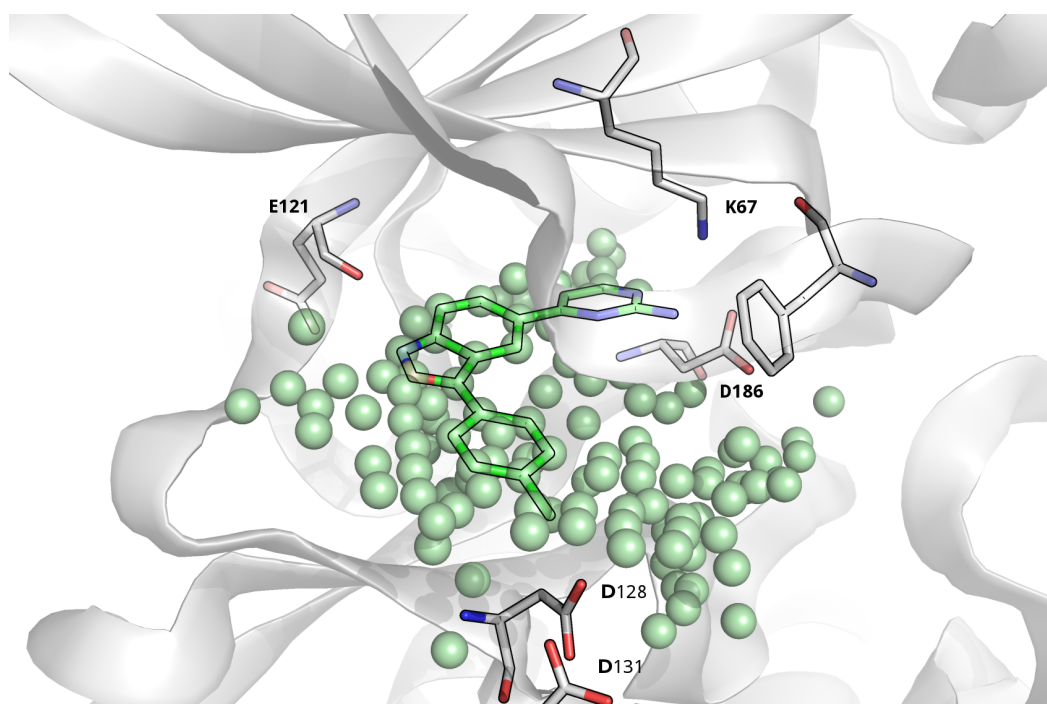


Figure 7.3: Prepared Pim-1 docking binding site for DOCK with matching spheres (green, sphere representation) with crystal ligand overlay (PDB:3BGP). Key residues (light grey, stick representation) are labelled.

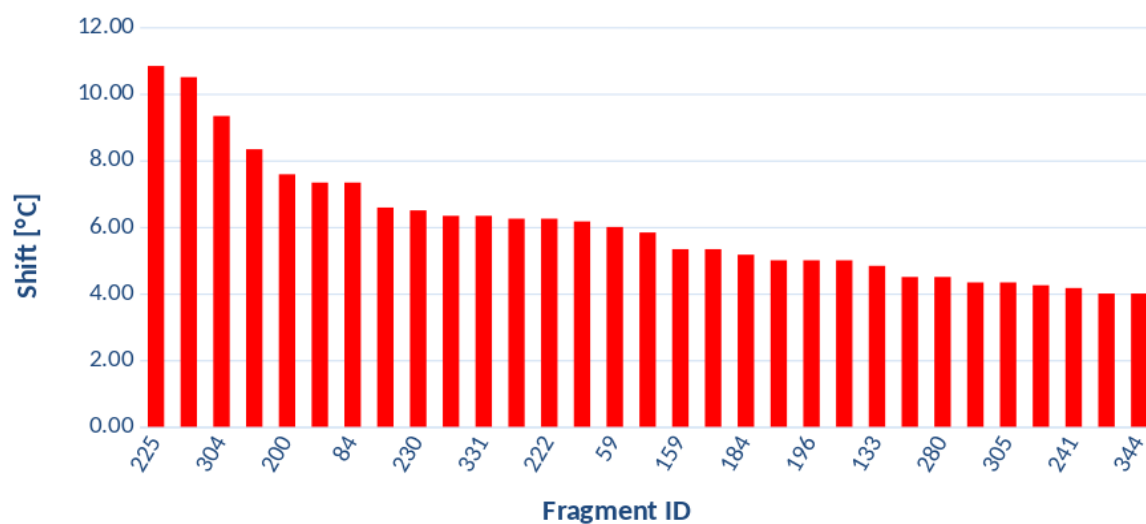


Figure 7.4: Pim-1 Thermal Shift Assay (TSA) results of 31 fragments that displayed a positive temperature shift of $\geq 4^\circ\text{C}$ (those with lower temperature shifts not shown) - fragment screen against internal fragment library

Section of protein	Residues	RMSD \bar{x} ($\sigma_{\bar{x}}$)
Binding pocket	L44-G55, V64-K67, E89, L93, I104, L120-E124, V126-D128, F130-D131, E171-L174, E180-F188	1.58 (0.18)
Catalytic lysine	K67	0.61 (0.18)
DFG loop	D186-F188	1.63 (0.20)
P-loop	G45-V52	0.99 (0.24)

Table 7.1: Mean RMSDs (\AA) of Pim-1 apo structure from MD simulations - Project 1

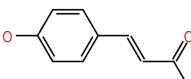
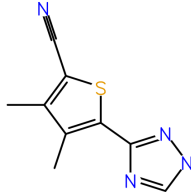
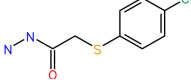
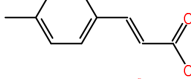
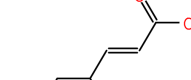
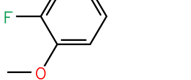
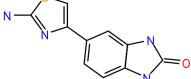
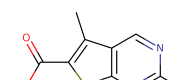
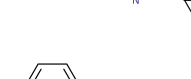
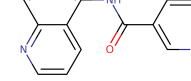
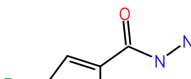
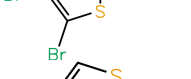
Frag	Predict binding mode? (Y/N) RMSD to best pose (Å)				PDB	
	FRED	HYBRID	SEED	DOCK		
1	N (>5Å)	N (>5Å)	Y (1.316)	N (>5Å)	5N4Z	
2	N (4.858)	N (>5Å)	Y (1.350)	N (>5Å)	5N4N	
3	Y (4.118)	N (4.616)	N (>5Å)	N (>5Å)	5N50	
4	Y (2.038)	Y (2.243)	Y (1.453)	N (>5Å)	5N4O	
5	Y (1.376)	Y (1.277)	Y (1.309)	N (>5Å)	-	
6	N (>5Å)	Y 3.087	Y (2.091)	Y (4.161Å)	5N4U	
7	Y (1.514)	Y 1.436	Y (1.566)	N (>5Å)	5N4V	
8	Y (1.410)	N (>5Å)	Y (2.882)	N (>5Å)	5MZL	
9	Y (2.558)	Y 3.129	Y (2.153)	N (>5Å)	5N4X	
10	N (3.789)	N (3.505)	Y (1.424)	N (>5Å)	5N51	
11	N (>5Å)	N (>5Å)	Y (1.301)	N (>5Å)	5N4Y	
12	Y (1.593Å)	Y 1.607	Y (1.613Å)	N (>5Å)	5N52	

Table 7.2: RMSD (Å) from the crystal structure of fragments from the internal fragment library as compared to the docked position for the 3BGP crystal structure for all docking codes used. 2D structure is also displayed - Project 1

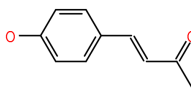
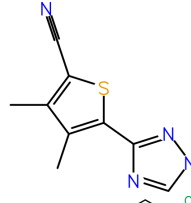
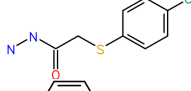
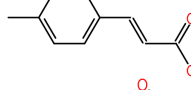
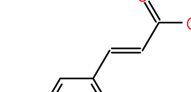
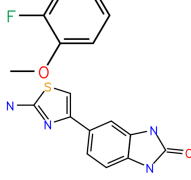
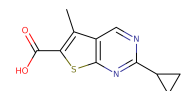
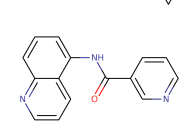
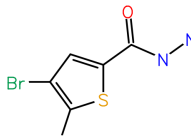
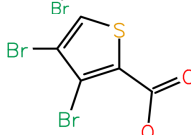
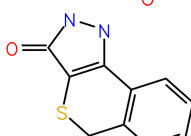
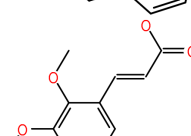
Predict binding mode? (Y/N) RMSD to best pose (Å)						
Frag	FRED	HYBRID	SEED	DOCK	PDB	
1	N (>5Å)	N (>5Å)	Y (1.889)	Y (1.889)	5N4Z	
2	Y (2.606)	Y (2.700)	Y (1.791)	Y (3.784)	5N4N	
3	Y (4.203)	Y (3.281)	Y (1.929)	Y (4.206)	5N50	
4	N (>5Å)	Y (2.966)	N (3.953)	N (4.750)	5N4O	
5	N (>5Å)	N (4.777)	Y (2.635)	N (>5Å)	-	
6	Y (2.664)	N (>5Å)	Y (2.706)	N (>5Å)	5N4U	
7	N (>5Å)	N (>5Å)	N (4.249)	N (>5Å)	5N4V	
8	N (4.787)	N (3.005)	Y (2.304)	N (>5Å)	5MZL	
9	N (>5Å)	N (3.791)	Y (2.244)	N (4.808)	5N4X	
10	Y (2.216)	Y (2.891)	Y (2.079)	N (4.446)	5N51	
11	Y (1.937)	Y (1.954)	Y (1.594)	-	5N4Y	
12	N (>5Å)	N (4.914)	N (4.086)	N (>5Å)	5N52	

Table 7.3: RMSD (Å) from the crystal structure of fragments from the internal fragment library as compared to the docked position for the apo_MD structure for all docking codes used. 2D structure is also displayed - Project 1

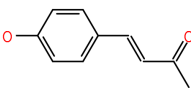
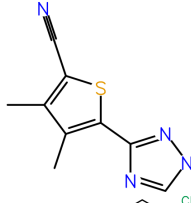
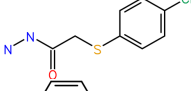
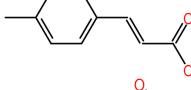
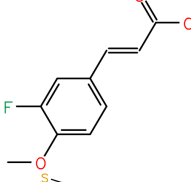
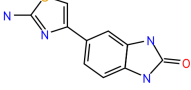
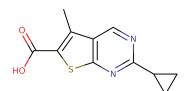
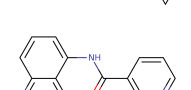
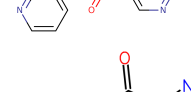
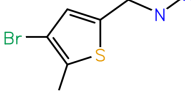
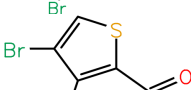
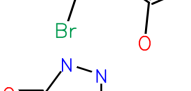
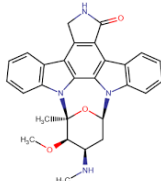
Predict binding mode? (Y/N) RMSD to best pose (Å)						
Frag	FRED	HYBRID	SEED	DOCK	PDB	
1	N (>5Å)	N (>5Å)	Y (2.559)	N (3.475)	5N4Z	
2	N (4.144)	N (>5Å)	Y (0.579)	N (>5Å)	5N4N	
3	N (>5Å)	N (>5Å)	N (3.987)	N (3.759)	5N50	
4	Y (1.586)	Y (1.057)	Y (2.602)	N (>5Å)	5N4O	
5	Y (2.492)	Y (2.527)	Y (1.343)	N (>5Å)	-	
6	N (>5Å)	N (>5Å)	Y (1.509)	N (>5Å)	5N4U	
7	Y (4.436)	Y (4.479)	Y (3.003)	Y (3.191)	5N4V	
8	N (>5Å)	N (>5Å)	Y (2.411)	N (>5Å)	5MZL	
9	N (4.414)	Y (3.048)	Y (1.834)	Y (4.414)	5N4X	
10	N (>5Å)	N (3.513)	Y (1.474)	N (4.708)	5N51	
11	N (3.594)	Y (0.666)	Y (0.528)	N (>5Å)	5N4Y	
12	Y (3.499)	Y (3.519)	Y (3.976)	N (>5Å)	5N52	

Table 7.4: RMSD (Å) from the crystal structure of fragments from the internal fragment library as compared to the docked position for the 3BGP_MD structure for all docking codes. 2D structure is also displayed - Project 1

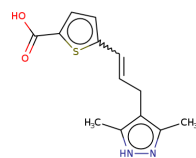
7.1.2 Project 2: synthetically-feasible extensions of an internal fragment library

Reaction	Number of protomers
1,2,4-triazole acetohydrazide	70
1,2,4-triazole carboxylic-acid or ester	23 759
benzimidazole derivatives carboxylic-acid or ester	710
benzothiophene	515
benzoxazole carboxylic-acid	120
Buchwald-Hartwig	29 704
Fischer indole	2
Friedlaender chinoline	1
Grignard alcohol	12 379
Grignard carbonyl	5612
Heck non-terminal vinyl	14
heteroaromatic nuc sub	719
Huisgen Cu-catalyzed 1,4-subst	64
Huisgen disubst-alkyne	110
Mitsunobu imide	30
Mitsunobu phenole	821
Mitsunobu sulfonamide	9
Mitsunobu tetrazole 1	35
Mitsunobu tetrazole 2	45
Mitsunobu tetrazole 3	3
Mitsunobu tetrazole 4	3
Negishi	14793
nucl sub aromatic ortho nitro	7
nucl sub aromatic para nitro	4
oxadiazole	9211
reductive amination	8198
Schotten-Baumann amide	190 351
Sonogashira	30
Stille	8962
sulfon amide	31
Suzuki	545
tetrazole connect regioisomere 1	875
tetrazole connect regioisomere 2	924
triaryl-imidazole	974
urea	135
Williamson ether	2394
Wittig	1140
Total	313 299

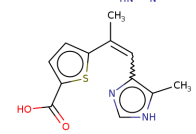
Table 7.5: Number of derivatives from fragments in Project 1 created *in silico* using PINGUI and reactions from Hartenfeller et al by reaction used - Project 2

Number	KolbLabID	K _i (μM)	2D Structure
Staurosporine -		4.71 μM	

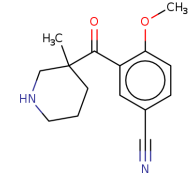
13 K002CT001



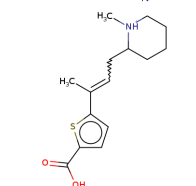
14 K002CT002



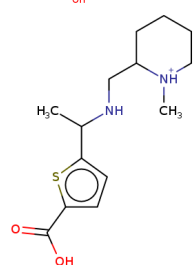
15 K002CT003



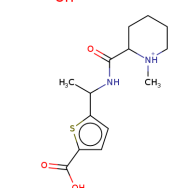
16 K002CT004



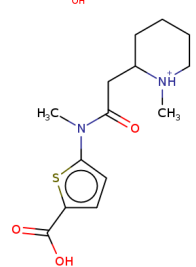
17 K002CT005



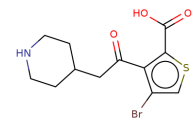
18 K002CT006



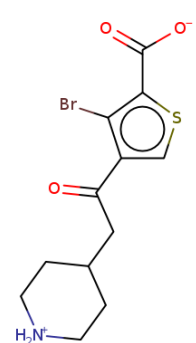
19 K002CT007



20 K002CT008

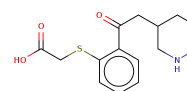


21 K002CT009



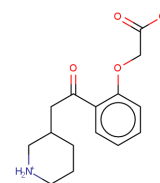
22

K002CT010



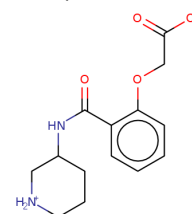
23

K002CT011



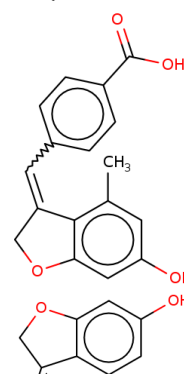
24

K002CT012

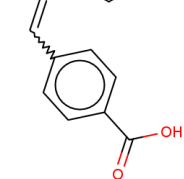


25

K002CT013

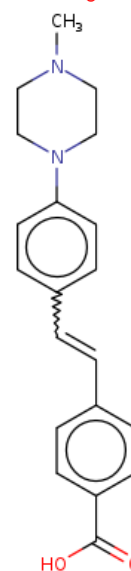


26

K002CT014 24.98 μ M

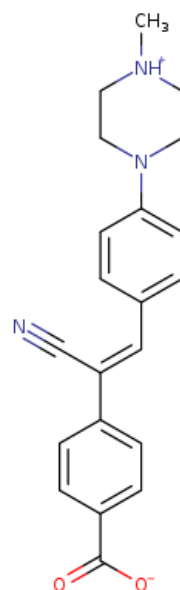
27

K002CT015 >10mM



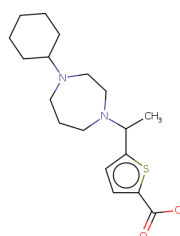
28

K002CT016



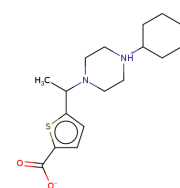
29

K002CT017 >10mM



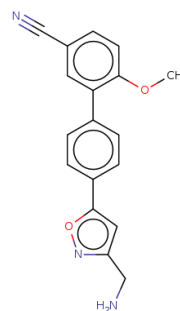
30

K002CT018 >10mM



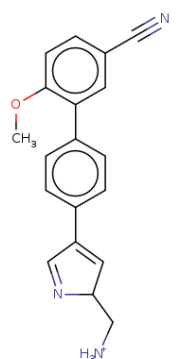
31

K002CT019



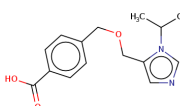
32

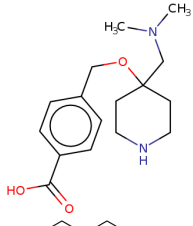
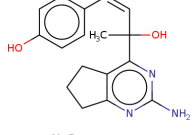
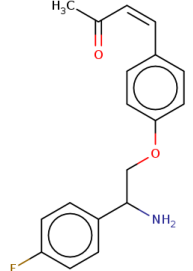
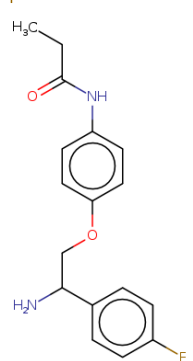
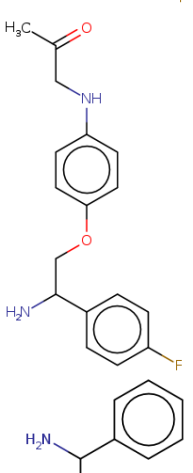
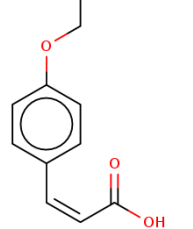
K002CT020

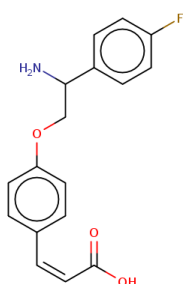
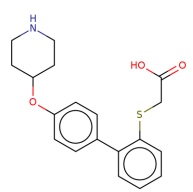
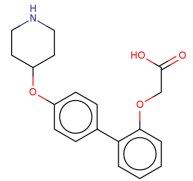
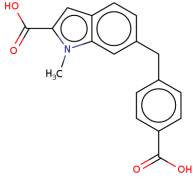
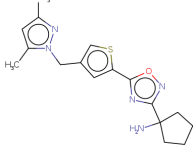
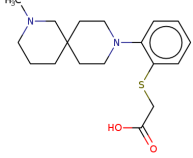
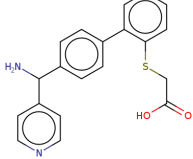
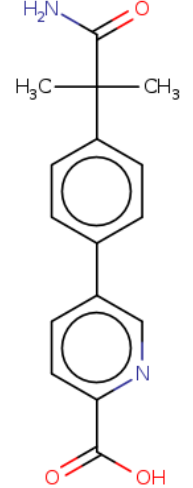


33

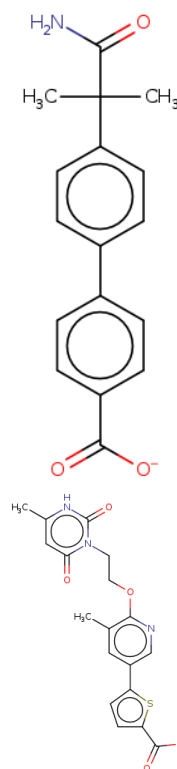
K002CT021



34	K002CT022	
35	K002CT023	
36	K002CT024	
37	K002CT025	
38	K002CT026	
39	K002CT027	141.14 μ M 

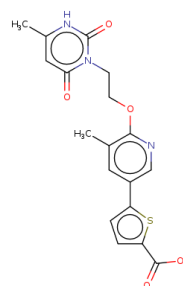
40	K002CT028	>10mM	
41	K002CT029		
42	K002CT030	91.47 μM	
43	K002CT031		
44	K002CT032		
45	K002CT033		
46	K002CT034		
47	K002CT035	191.76 μM	

48

K002CT036 250.44 μ M

49

K002CT037

**Table 7.6:** Complete list of predicted Pim-1 inhibitors, K_i of all that were tested and 2D structure - Project 2

	26	27	29
Overall MMFF energy terms			
MMFF VdW	173.10853	174.99911	211.79359
MMFF Coulomb	-773.74868	-1054.68906	-1010.24262
MMFF Torsion	73.51171	92.36242	86.00307
Protein shell MMFF terms:			
MMFF VdW	163.55022	155.04904	179.39431
MMFF Coulomb	-634.13971	-834.09671	-847.24709
MMFF Torsion	69.88636	73.76472	89.46652
Protein shell - ligand MMFF terms			
MMFF VdW	-11.24085	-7.97985	-4.64507
MMFF Coulomb	-143.96872	-172.25685	-140.44702
Ligand MMFF Intramol. Terms			
MMFF VdW	20.79916	27.92992	37.04436
MMFF Coulomb	4.35975	-48.33549	-22.54851
MMFF Bond	0.0000	0.0000	0.0000
MMFF Bend	0.0000	0.0000	0.0000
MMFF StretchBend	0.0000	0.0000	0.0000
MMFF Torsion	3.62536	18.5977	-3.46346
MMFF Improper Torsion	0.0000	0.0000	0.0000
Ligand MMFF Intramol. Energy	28.78427	-1.80788	11.03239
Overall Ligand-Protein Interaction terms			
-- VdW	-15.24877	-13.79325	-8.79115
-- Coulomb diel=1.0	157.15608	-163.84875	-135.23599
-- Protein desolv (PB)	11.87588	18.05575	16.61079
-- Ligand desolv (PB)	18.79844	19.47878	17.90345
-- Solvent screening (PB)	-177.23943	129.67464	104.97469
Overall Lig-Prot Interaction	-4.65779	-10.43284	-4.53821
Total energy	24.12647 kcal/mol	-12.24072 kcal/mol	6.49418 kcal/mol

Table 7.7: Selected Szybki minimisation output for synthesised inhibitors **26**, **27** and **29** - Project 2

	30	39	40
Overall MMFF energy terms			
MMFF VdW	178.32713	173.22691	178.75834
MMFF Coulomb	-994.08532	-849.55585	-953.7639
MMFF Torsion	74.43372	75.55722	77.25445
Protein shell MMFF terms:			
MMFF VdW	152.00414	149.60911	158.84842
MMFF Coulomb	-793.83878	-633.62987	-776.76304
MMFF Torsion	75.79929	66.65698	70.88795
Protein shell - ligand MMFF terms			
MMFF VdW	-12.5733	1.3277	-4.06612
MMFF Coulomb	-183.92707	-171.73928	-132.58372
Ligand MMFF Intramol. Terms			
MMFF VdW	38.89629	22.2901	23.97603
MMFF Coulomb	-16.31947	-44.1867	-44.41714
MMFF Bond	0.0000	0.0000	0.0000
MMFF Bend	0.0000	0.0000	0.0000
MMFF StretchBend	0.0000	0.0000	0.0000
MMFF Torsion	-1.3656	8.9002	6.3665
MMFF Improper Torsion	0.0000	0.0000	0.0000
Ligand MMFF Intramol. Energy	21.21125	-12.99636	-14.07461
Overall Ligand-Protein Interaction terms			
-- VdW	-16.18952	-3.71935	-8.58616
-- Coulomb diel=1.0	-162.83585	-222.80383	-194.34521
-- Protein desolv (PB)	20.79349	17.2907	19.59708
-- Ligand desolv (PB)	19.83262	24.08791	18.44216
-- Solvent screening (PB)	122.8871	170.24762	151.08281
Overall Lig-Prot Interaction	-15.51216	-14.89695	-13.80933
Total energy	5.69909 kcal/mol	-27.89331 kcal/mol	-27.88393 kcal/mol

Table 7.8: Selected Szybki minimisation output for synthesised inhibitors **30**, **39** and **40** - Project 2

	42	47	48
Overall MMFF energy terms			
MMFF VdW	196.33979	199.14136	194.1482
MMFF Coulomb	-1131.72547	-942.02649	-866.95849
MMFF Torsion	119.45029	86.9706	83.10517
Protein shell MMFF terms:			
MMFF VdW	184.86334	175.19889	169.06271
MMFF Coulomb	-905.12258	-925.08184	-743.19743
MMFF Torsion	101.90064	76.78063	73.05085
Protein shell - ligand MMFF terms			
MMFF VdW	-7.2298	-4.80154	-11.25881
MMFF Coulomb	-195.32913	-86.67719	-131.8179
Ligand MMFF Intramol. Terms			
MMFF VdW	18.70625	28.74401	36.3443
MMFF Coulomb	-31.27376	69.73254	8.05684
MMFF Bond	0.0000	0.0000	0.0000
MMFF Bend	0.0000	0.0000	0.0000
MMFF StretchBend	0.0000	0.0000	0.0000
MMFF Torsion	17.54965	10.18997	10.0543
MMFF Improper Torsion	0.0000	0.0000	0.0000
Ligand MMFF Intramol. Energy	4.98214	108.66652	54.45545
Overall Ligand-Protein Interaction terms			
-- VdW	-10.34907	-10.1994	-15.50657
-- Coulomb diel=1.0	-184.8342	125.85612	136.16232
-- Protein desolv (PB)	19.76263	16.49491	13.70759
-- Ligand desolv (PB)	20.32186	19.62251	19.44108
-- Solvent screening (PB)	145.88319	-152.89174	-159.19449
Overall Lig-Prot Interaction	-9.21559	-1.1176	-5.39006
Total energy	-4.23345 kcal/mol	107.54893 kcal/mol	49.06539 kcal/mol

Table 7.9: Selected Szybki minimisation output for synthesised inhibitors **42**, **47** and **48** - Project 2

Number	InternalID	K_i (μ M)	2D Structure
50	LHA039	>10mM	
51	LHA047	>10mM	
52	LHA085	132.11 μ M	
53	LHA086	>10mM	
54	LHA100	>10mM	
55	LHA119	>10mM	

Table 7.10: Predicted Pim-1 inhibitors, aniline triazole series - Project 3

7.1.3 Project 3: novel inhibitors from the ZINC dataset

ID	Kinase tested										
Compound	Abl	ASK1	B-Raf	Lyn	MAPK2	mTOR	Pim-1	Pim-2	Plk4	Rsk2	TTBK2
56	111	100	100	102	97	99	98	118	104	104	103
57	82	94	105	86	92	100	77	95	98	96	104
58	88	100	94	96	101	72	44	57	96	97	105
59	105	105	92	100	111	102	83	105	112	96	107
60	104	110	101	95	101	101	94	92	107	87	99
61	96	99	110	96	98	101	99	109	106	84	99

Table 7.11: Mean % residual activity at 10 μ M of Pim-1 inhibitors as determined by third-party vendor (Eurofins) using method in [97] - Project 3

Compound	In-house IC ₅₀ (μ M)	Eurofins IC ₅₀ (μ M)
57	308.02	56.5
58	13.2	4.2
59	NOT TESTED	> 100 μ M
61	1843	> 100 μ M

Table 7.12: Calculated IC₅₀'s of Pim-1 inhibitors as determined by third-party vendor (Eurofins) using method in [97] - Project 3

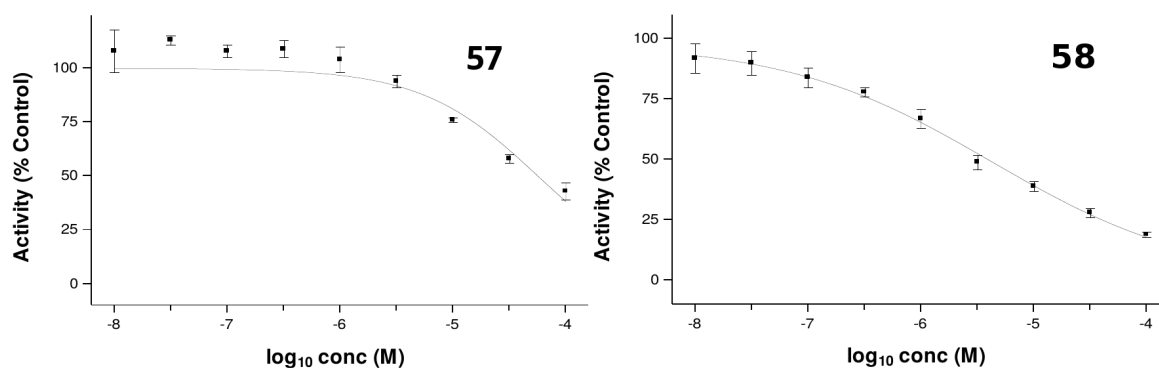


Figure 7.5: IC₅₀ curves for compounds **57** and **58** as determined by third-party vendor (Eurofins) using method in [97] (Eurofins)

7.1.4 Project 4: extensions of truncated fragment f200

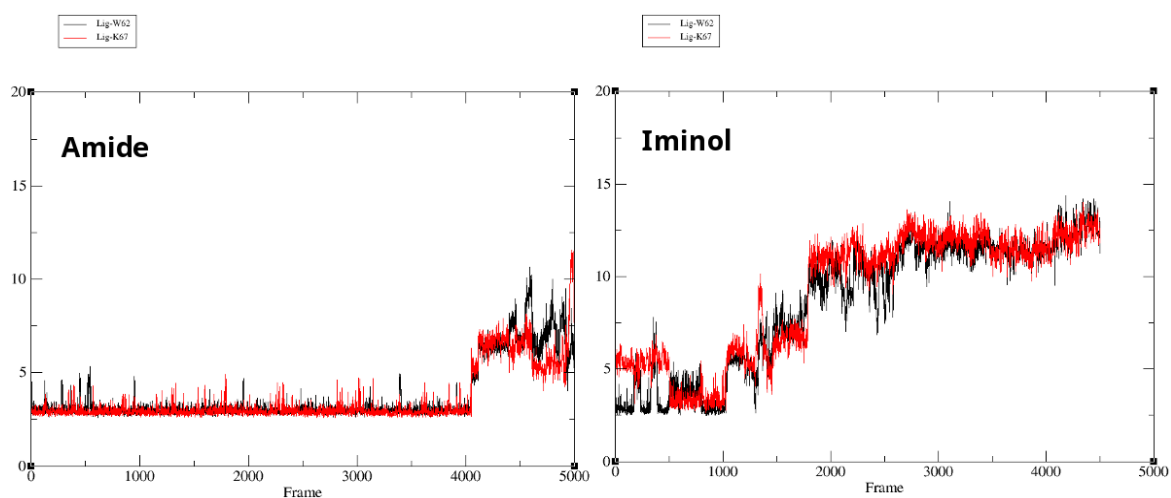


Figure 7.6: Mean distances (\AA) between simulated amide (left) and iminol (right) tautomers with W62 (black trace) and K67 (red trace) against Pim-1 for first 10ns of simulations

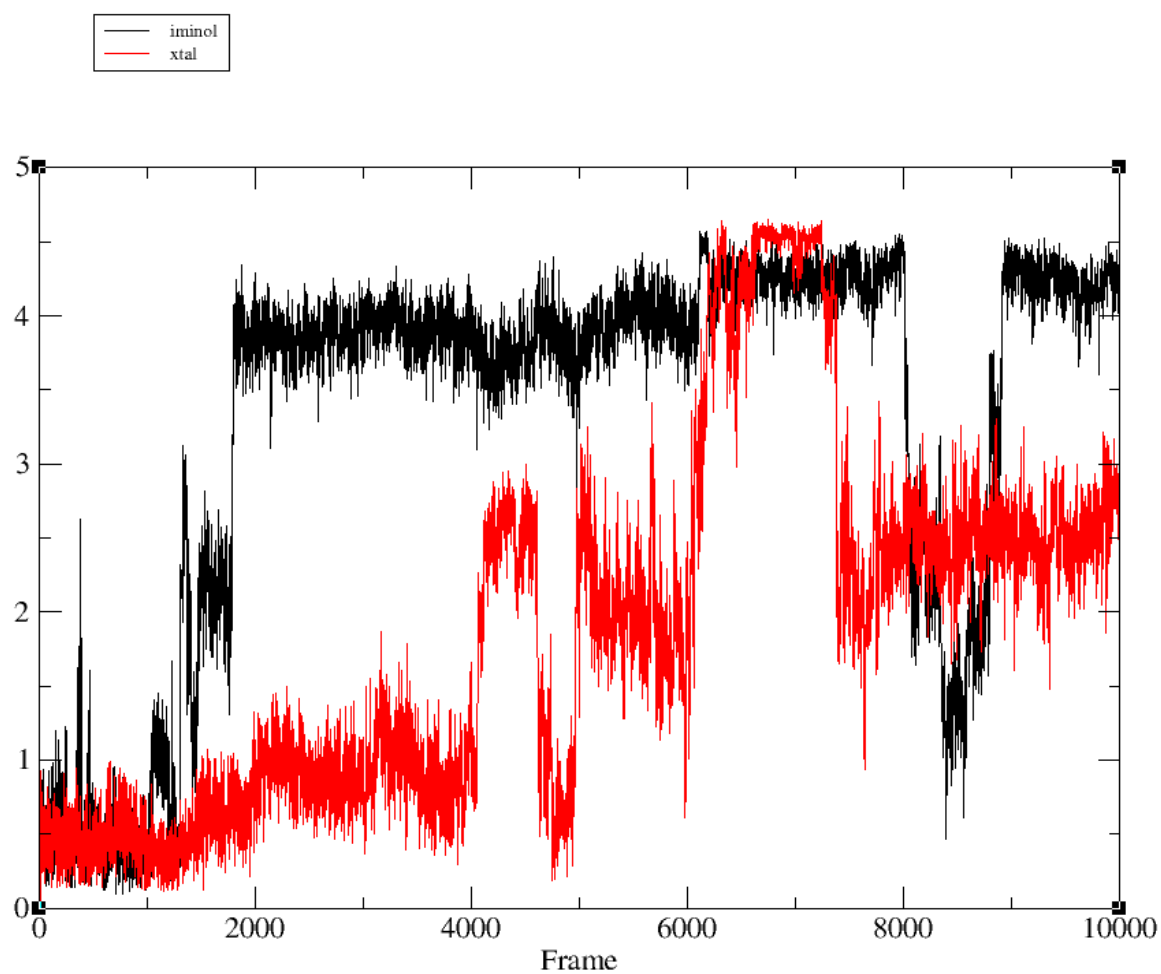


Figure 7.7: Mean RMSD (Å) between simulated amide (red trace) and iminol (black trace) tautomers against Pim-1 for first 20ns of simulations

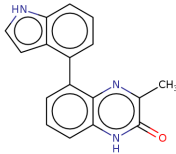
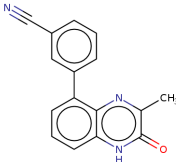
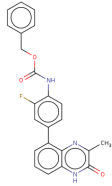
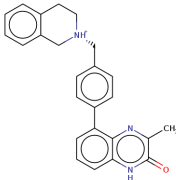
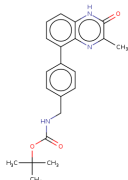
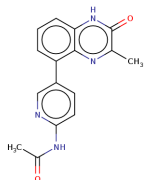
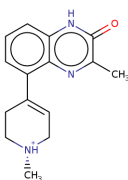
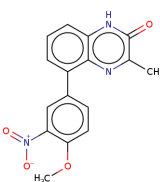
Number	KolblabID	2D Structure
62	K004CT001	
63	K004CT002	
64	K004CT003	
65	K004CT004	
66	K004CT005	
67	K004CT006	
68	K004CT007	
69	K004CT008	

Table 7.13: Predicted quinoxaline derivative inhibitors, unique IDs and 2D structures - Project 4

7.2 Novel inhibitors against the Smoothed receptor

Compounds	Assay kit ingredients
Vitamin K	Pierce BCA assay kit (Thermo-Fisher)
Vitamin E	Power SYBR Green Cells-To-CT (Thermo-Fisher)
PEG3	SigmaFast EDTA-free protease inhibitor cocktail
Ceramide	
DDM	
Monoolein	

Table 7.14: Exhaustive list of molecules ruled out by collaborators as fitting within the CURED

Residue number	Unpublished structure	PDB code	
		5L7I	4N4W
H63	HIP	HIE	HIE
H103	HID	-	-
H227	HID	HID	HIE
H231	HIE	HIE	HIP
H340	HID	HIP	HIE
H361	HIE	HIE	HIE
H433	-	HIP	HIP
H470	HIE	HIE	HIE
H1078	HIP		-

Table 7.15: Protonation states of Histidines in all Smoothed structures used for docking where HID=only the δ Nitrogen on imidazole sidechains were protonated, HIE= ϵ Nitrogen is protonated and HIP=both Nitrogens are protonated.

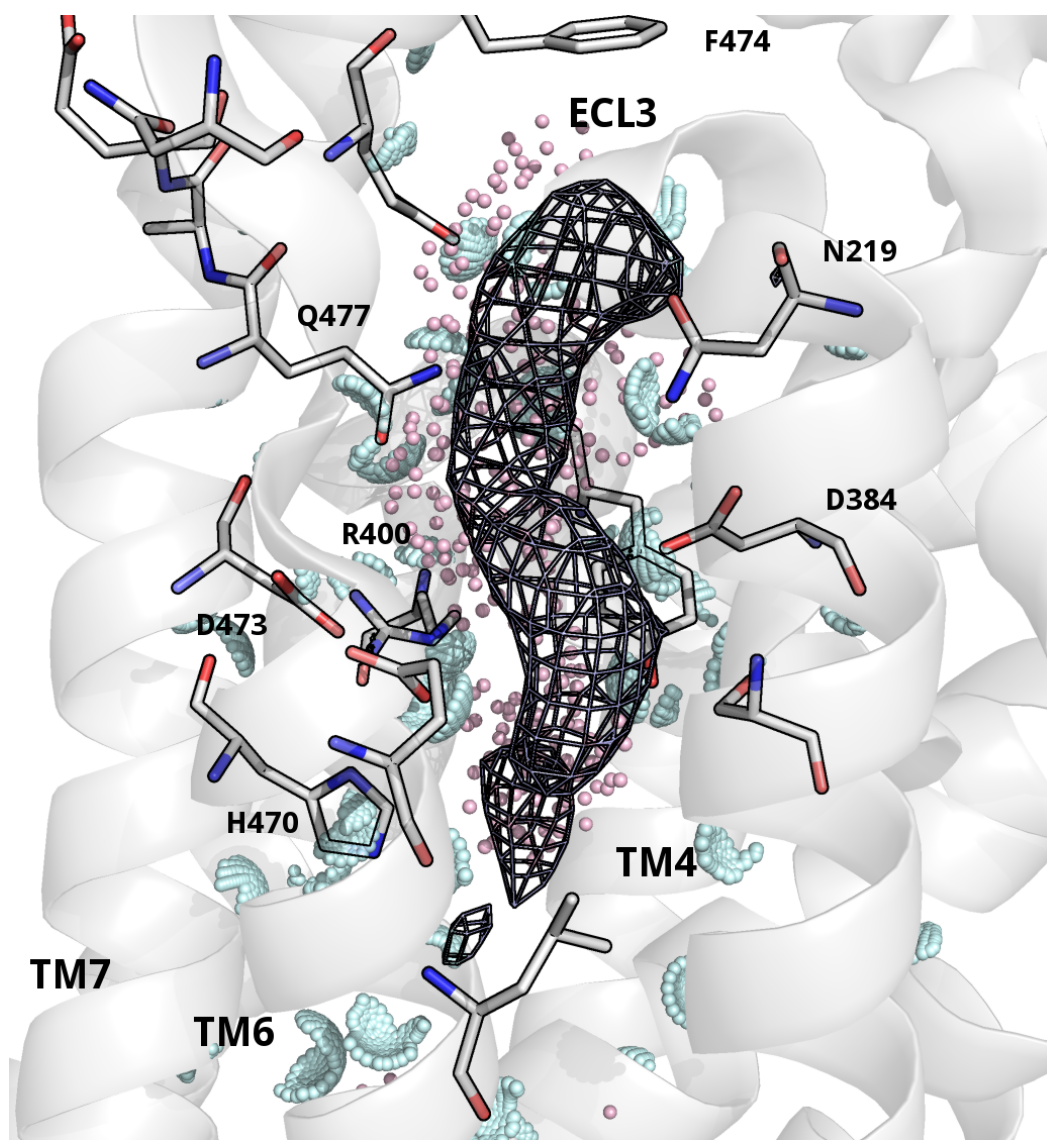


Figure 7.8: Prepared SMO docking binding site polar (light blue, small spheres) and apolar (light pink, small spheres) vectors for SEED with CURED Fo-Fc density for reference, key residues (light grey, stick representation) are labelled - unpublished X-ray structure

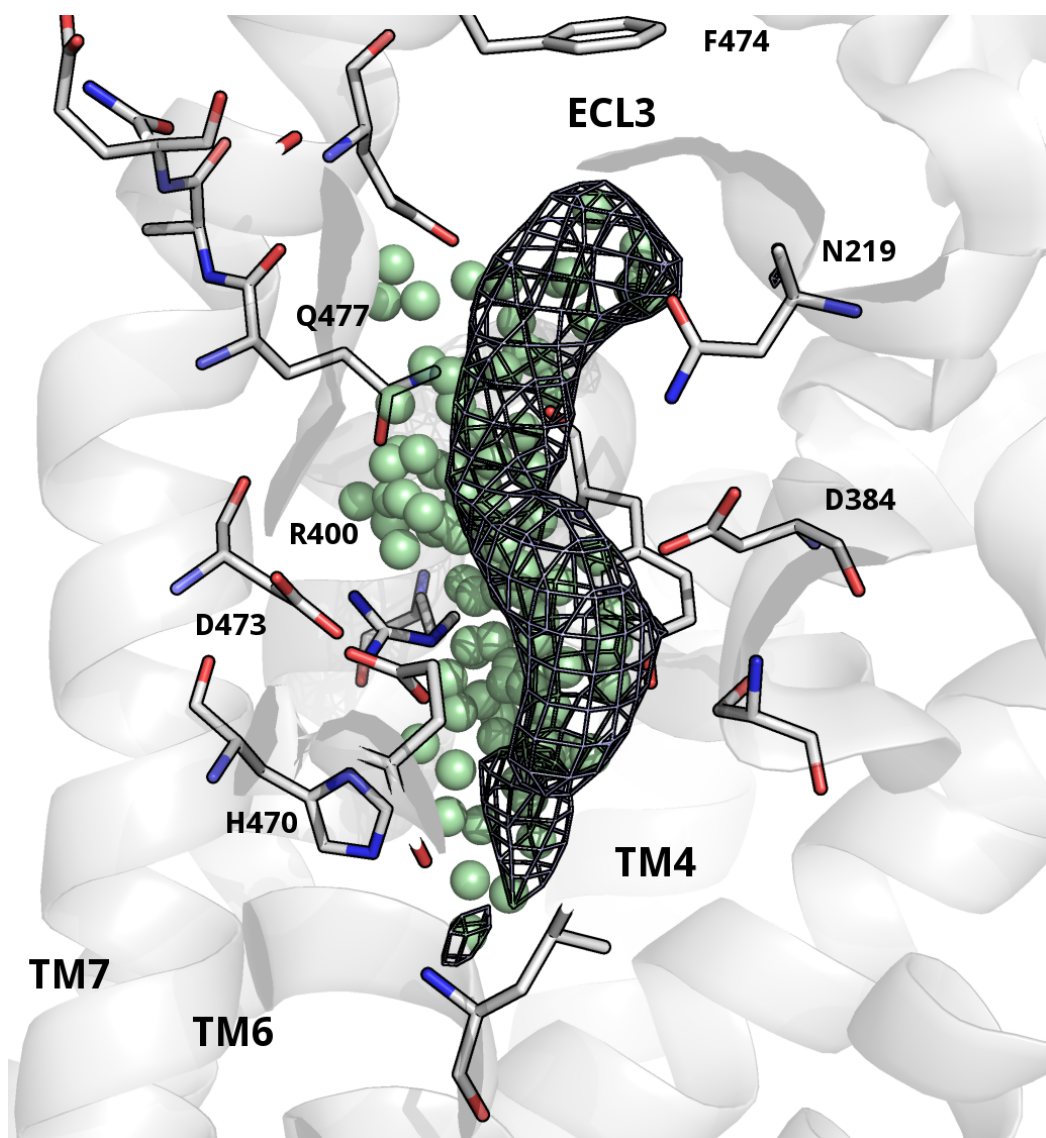


Figure 7.9: Prepared SMO docking binding site for DOCK with matching spheres (green, sphere representation) with CURED Fo-Fc density for reference, key residues (light grey, stick representation) are labelled - unpublished X-ray structure

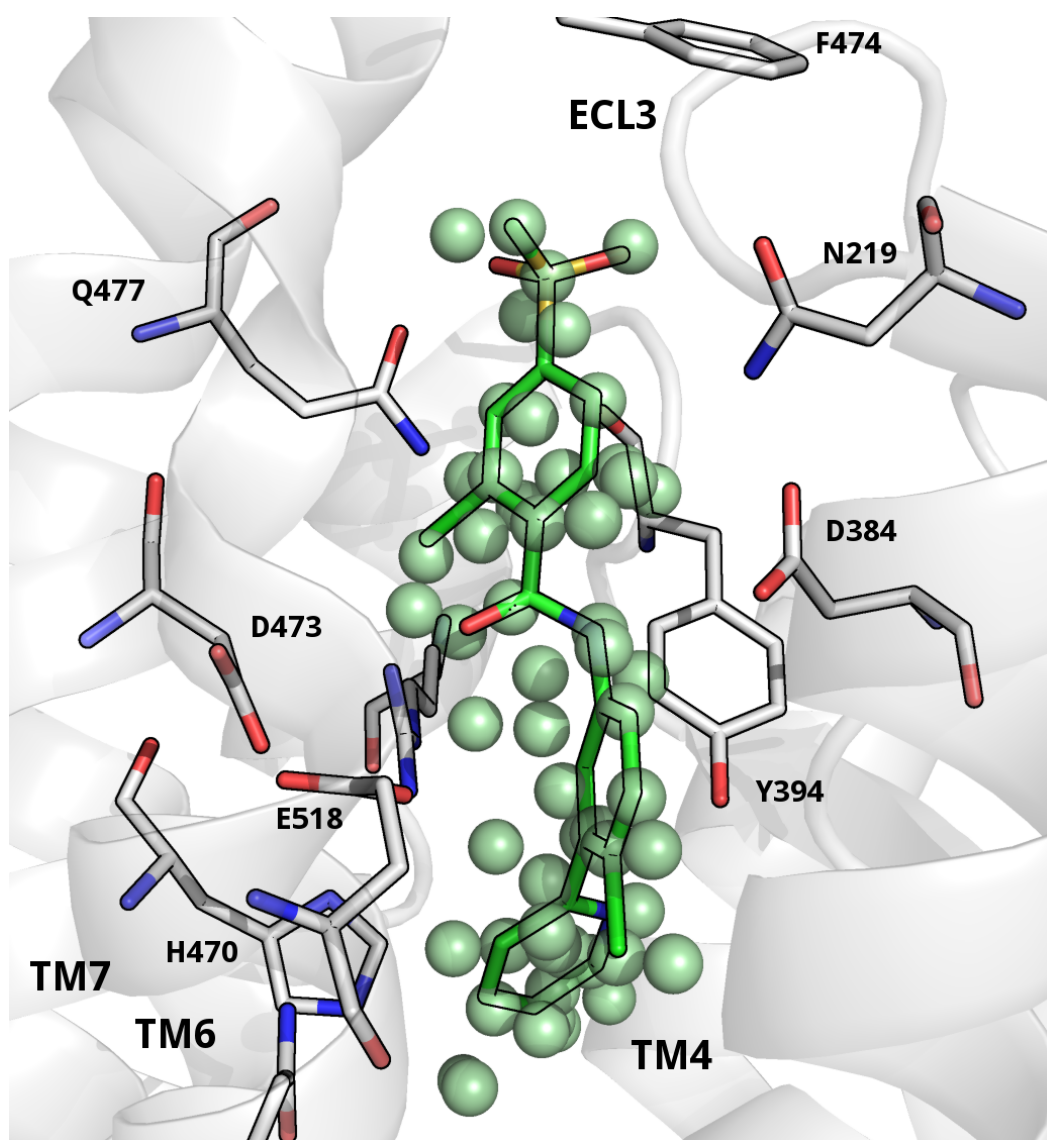


Figure 7.10: Prepared SMO docking binding site for DOCK with matching spheres (green, sphere representation) with Vismodegib (green, stick representation) for reference, key residues (light grey, stick representation) are labelled - PDB: 5L7I

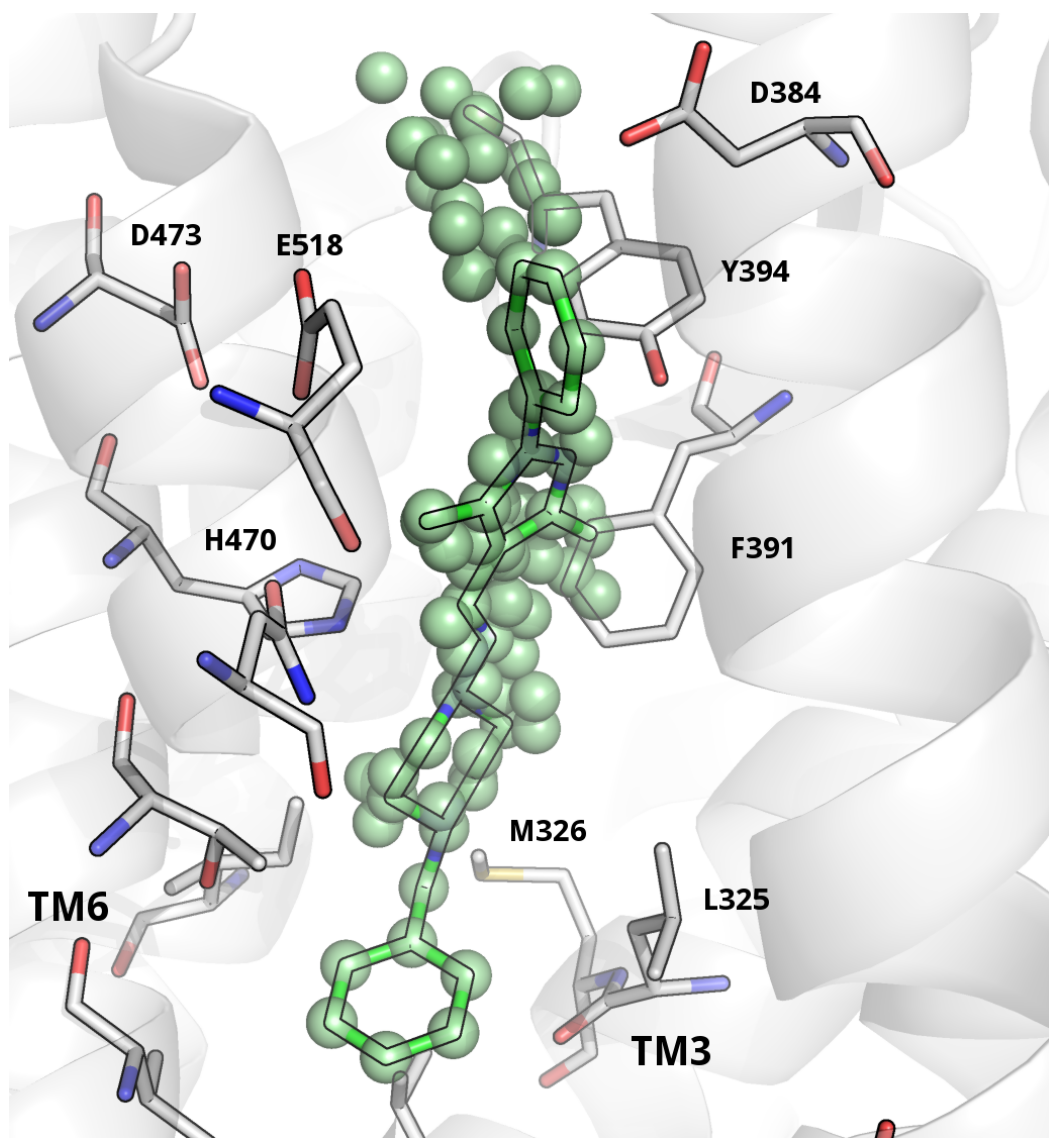


Figure 7.11: Prepared SMO docking binding site for DOCK with matching spheres (green, sphere representation) with SANT-1 (green, stick representation) for reference, key residues (light grey, stick representation) are labelled - PDB: 4N4W

1, 4-dimethylbenzene	Cytosine	N-methylbenzimidazole
1,2-difluorobenzene	D-Proline	N-Methylindole
1,2-dimethylbenzene	Decahydroquinoline	N-Methylpyrrole
1,3-dimethylbenzene	Decaline	N-pyrrolidinaldehyde-2-carboxylic acid
2-aminopyradine	Delta-valerolactam	N, N, Dimethylaniline
2-aminopyrimidine-4-one	Dibenzocyclohexane	N(3)-methyl-2-aminopyrimidinone
2-methyl-2-butene	Dihydroxy-benzene	Napthalene
2-methyl-2-phenylpropane	Dihydroxy-cyclohexane	Nitrobenzene
2-methylthiophene	Dimethylammonium	Phenole
2-phenylethanol	Dimethylether	Phenylethane
2-Pyrrolidinone	Dinitrobenzene	Phenylsulfonamide
2,2-Dimethylpropane	Diphenyl Ether	Phenylurea
2,3-Dihydroxy-tetrahydrofuran	Ethane	Phosphate anion
2,3,4-furantricarboxylic-acid	Ethanol	Piperadine
2,4-diaminopyrimidine	Ethene	Piperazine
2,4-dichlorotoluene	Ethylammonium	Proline
2,5-Diketopiperazine	Fluorobenzene	Propane
3-methoxyanisole	Guanine	Propene
3-methylanisole	Quinoline-4,5, dihydroxyl-1-amine	Propionic acid
3,6-dimethyl-2,5-diketo-1,4-piperazine	Quinoline-4,7-diamine	Purine
5-amidine indole	Quinoline-3,10-diamine	Pyridine
5-methylisooxazole	Quinoline-3-E-methylamide-10-amine	Pyrrole
5-phenyl-1,4-benzodiazepine	Quinoline-3-Z-methylamide-10-amine	Quinazoline
5-phenyl-1,4-benzodiazepine-2-one	Imidazole – less polar	Quinoline
Acetate	Imidazole – more polar	Tert-butane
Acetone	Indane	Tetrahydro-2-pyrimidinone
Acetophenone	Indole	Tetrahydro-quinoline
Adamantane	Isopropanol	Tetrahydrofuran
Adenine	Isoquinazoline	Tetrahydropyrrole
alpha-Carboline	Isoquinoline	Tetraline
Aminomethane	L-Proline	Thiazole
Aniline	Meso-Inositol	Thiazolidinone
Anisol	Methane	Thiophene
Benzamide	Methanol	Toluene
Benzamidine	Methanol	Trans-Dihydroxy-cyclohexane (anti)
Benzene	Methyl benzoate	Trifluoroethane
Benzodioxolane	Methyl guanidine	Trifluoromethane
Benzoic acid	Methyl sulfone	Trifluoromethylbenzene
Benzonitrile	Methylamidine	Trimethylammonium
Benzothiophene	Methylammonium	Uracil
Beta-carboline	Methylcyclopropane	Water
Bromobenzene	Methylphenol	
Butene	Methylphenylsulfide	
Chlorobenzene	Methylphenylsulfone	
cis-(D-Ala,D-Ala)-2,5-Diketopiperazine	Methylphenylsulfoxide	
Cis-Dihydroxycyclohexane	N-ethylimidazole	
Cyclohexane	N-ethylpyrrolidinone	
Cyclohexoguanidine	N-Methyl Methylamide	
Cyclopentanone	N-Methyl-methylsulfonamide	
Cyclopropane	N-Methylacetamide	

Table 7.16: All molecular 'probes' used in docking to the unpublished structure to identify the molecule in the CURED.

Fragment	Interaction energy: Max($\mu \pm \sigma$)	Notes
Highly polar molecules		
Benzoic acid	No favorable poses	Apolar but acid moiety
Methyl sulfone	No favorable poses	Small strongly polar molecule
Imidazole – more polar	No favorable poses	Small ring, few seeded positions
Propionic acid	No favorable poses	Small acid
Phosphate anion	No favorable poses	
Small, polar molecules		
Acetate	-0.8	One pose
Acetone	-8.25 (-3.99 \pm 1.43)	
Aminomethane	-7.22 (-3.62 \pm 1.22)	
D-Proline	-5.69 (-1.98 \pm 1.64)	
L-Proline	-13.39 (-2.50 \pm 2.60)	
Methanol	-6.83 (-2.46 \pm 1.30)	
Trifluoroethane	-5.13 (-1.88 \pm 1.14)	
Thiophene	-6.39 (-2.77 \pm 1.22)	
Dimethylammonium	-14.26 (-4.32 \pm 2.85)	Polar interactions outside blob, large receptor desolvation costs
Dimethylether	-8.06 (-3.29 \pm 1.38)	No convincing poses, outside blob
Ethylammonium	-14.18 (-4.21 \pm 3.56)	Polar interactions outside blob, large receptor desolvation costs
Ethanol	-8.13 (-3.32 \pm 1.68)	Polar interactions outside blob, large receptor desolvation costs
Methylamidine	-18.85 (-5.75 \pm 3.55)	
Methylammonium	-14.43 (-3.04 \pm 3.09)	Polar interactions outside blob, large receptor desolvation costs
Methyl guanidine	-17.79 (-6.26 \pm 3.44)	Polar interactions outside blob, large receptor desolvation costs
N-Methyl-methylsulfonamide	-17.79 (-2.69 \pm 1.77)	Polar interactions outside blob, large receptor desolvation costs
N-Methyl Methylamide	-10.34 (-3.48 \pm 1.99)	
N-Methylacetamide	-10.33 (-4.04 \pm 2.00)	
Pyrrole	-6.57 (-3.15 \pm 1.45)	
Tetrahydropyrrole	-16.41 (-5.83 \pm 4.24)	Scores inflated by polar interactions that are outside blob.
Tetrahydrofuran	-9.52 (-5.06 \pm 1.81)	
Thiazole	-5.84 (-2.62 \pm 0.97)	
Thiazolidinone	-9.93 (-3.47 \pm 1.45)	Polar interactions outside blob, large receptor desolvation costs
Trifluoromethane	-5.13 (-1.88 \pm 1.14)	
Trimethylammonium	-8.69 (-3.44 \pm 1.83)	
Water	-2.97 (-1.49 \pm 0.72)	
Small, apolar molecules		
Ethane	-7.32 (-2.61 \pm 1.36)	
Ethene	-5.41 (-2.14 \pm 1.04)	
Methane	-4.30 (-2.49 \pm 0.91)	
Propane	-8.36 (-3.71 \pm 1.40)	
Propene	-7.45 (-3.14 \pm 1.12)	
Tert-butane	-8.03 (-3.90 \pm 1.70)	
2,2-Dimethylpropane	-8.04 (-4.67 \pm 1.81)	

Table 7.17: Calculated interaction energies in kcal·mol⁻¹, (maximum, mean and standard deviation) of molecular probes where poses were found outside CURED.

Fragment	Interaction Max($\mu \pm \sigma$)	energy:	Fragment	Interaction Max($\mu \pm \sigma$)	energy:
Dibenzocyclohexane	-9.14 (-4.34 \pm 2.60)		Hag 5	-12.24 (-5.51 \pm 3.41)	
Quinoline-4,5, dihydroxyl-1-amine	-13.33 (-4.62 \pm 3.27)		Indane	-10.54 (-5.04 \pm 2.88)	
alpha-Carboline	-12.48 (-4.94 \pm 3.24)		Indole	-9.45 (-4.25 \pm 2.45)	
Acetphenone	-11.83 (-4.76 \pm 2.66)		Isopropanol	-9.09 (-4.53 \pm 1.70)	
Adamantane	-11.83 (-10.33 \pm 0.515)		Isoquinoline	-9.50 (-4.28 \pm 2.47)	
Adenine	-9.5 (-3.62 \pm 2.00)		Isoquinazoline	-9.68 (-4.05 \pm 2.64)	
5-amidine indole	-25.43 (-6.67 \pm 5.35)		2-methyl-2-butene	-8.90 (-4.43 \pm 1.87)	
2-aminopyradine	-7.95 (-3.29 \pm 1.72)		Methyl benzoate	-11.72 (-4.64 \pm 2.93)	
2-aminopyrimidine-4-one	-10.58 (-4.19 \pm 1.88)		Methylcyclopropane	-7.98 (-4.59 \pm 1.21)	
Aniline	-10.37 (-5.20 \pm 2.08)		1,3-dimethylbenzene	-9.87 (-4.58 \pm 2.48)	
Anisol	-10.33 (-4.19 \pm 2.36)		N(3)-methyl-2-aminopyrimidinone	-11.58 (-4.64 \pm 2.41)	
Imidazole – less polar version	-8.22 (-3.45 \pm 1.44)		3-methylanisole	-11.40 (-4.46 \pm 2.89)	
Beta-carboline	-8.22 (-4.41 \pm 3.08)		N-methylbenzoimidazole	-11.27 (-4.64 \pm 2.62)	
Benzamidine	-22.43 (-5.57 \pm 4.46)		5-methylisooxazole	-9.86 (-3.89 \pm 1.85)	
Benzene	-6.50 (-2.76 \pm 1.76)		2-methyl-2-phenylpropane	-11.43 (-5.56 \pm 3.17)	
Bromobenzene	-10.23 (-4.73 \pm 2.58)		Methylphenylsulfide	-9.40 (-3.69 \pm 2.37)	
Butene	-8.07 (-4.56 \pm 1.05)		Methylphenylsulfone	-11.32 (-3.31 \pm 2.36)	
Benzamide	-11.38 (-4.17 \pm 2.47)		Methylphenylsulfoxide	-10.00 (-3.85 \pm 2.39)	
Benzodioxolane	-9.56 (-3.76 \pm 2.47)		2-methylthiophene	-7.87 (-3.46 \pm 1.65)	
Benzonitrile	-8.34 (-3.01 \pm 1.92)		Methylphenol	-9.29 (-4.17 \pm 2.29)	
Benzothiophene	-8.69 (-3.68 \pm 2.32)		3-methoxyanisole	-12.74 (-4.86 \pm 3.14)	
Cis-Dihydroxycyclohexane	-13.03 (-5.20 \pm 3.04)		Meso-Inositol	-13.65 (-5.19 \pm 3.03)	
Cyclohexoguanidine	-21.30 (-6.48 \pm 4.39)		Napthalene	-10.14 (-4.39 \pm 2.87)	
Chlorobenzene	-9.51 (-4.16 \pm 2.67)		Nitrobenzene	-10.30 (-4.28 \pm 2.45)	
Cyclohexane	-8.78 (-4.09 \pm 2.65)		N-Methylindole	-10.53 (-4.22 \pm 2.52)	
Cyclopentanone	-8.22 (-4.66 \pm 1.53)		N-Methylpyrrole	-8.09 (-3.83 \pm 1.74)	
Cyclopropane	-6.57 (-2.92 \pm 1.40)		1,2-dimethylbenzene	-9.51 (-4.41 \pm 2.64)	
Cytosine	-9.81 (-4.25 \pm 1.99)		5-phenyl-1,4-benzodiazepine-2-one	-16.43 (-5.51 \pm 4.18)	
Decahydroquinoline	-13.52 (-6.41 \pm 3.67)		5-phenyl-1,4-benzodiazepine	-13.27 (-4.62 \pm 2.53)	
Decaline	-12.90 (-5.78 \pm 3.56)		1, 4-dimethylbenzene	-10.48 (-4.48 \pm 2.76)	
Delta-valerolactam	-10.22 (-5.54 \pm 1.96)		Phenole	-9.65 (-3.92 \pm 2.08)	
2,3-Dihydroxy-tetrahydrofuran	-12.48 (-5.45 \pm 2.35)		Phenylethane	-10.45 (-4.15 \pm 2.55)	
Dihydroxy-benzene	-8.36 (-4.23 \pm 2.04)		2-phenylethanol	-10.70 (-3.75 \pm 2.53)	
Dihydroxy-cyclohexane	-13.03 (-7.24 \pm 2.64)		Phenylsulfonamide	-11.65 (-3.25 \pm 2.44)	
2,4-diaminopyrimidine	-8.90 (-3.80 \pm 1.90)		Phenylurea	-12.50 (-4.29 \pm 2.81)	
2,4-dichlorotoluene	-12.24 (-5.00 \pm 3.21)		Piperazine	-16.07 (-6.75 \pm 4.99)	
1,2-difluorobenzene	-8.75 (-3.68 \pm 2.10)		Piperadine	-18.01 (-5.83 \pm 4.62)	
N, N, Dimethylaniline	-11.58 (-4.15 \pm 2.78)		Proline	-13.87 (-2.51 \pm 2.48)	
Dinitrobenzene	-12.38 (-4.85 \pm 3.09)		Purine	-10.51 (-3.45 \pm 2.04)	
Diphenyl Ether	-10.87 (-3.39 \pm 2.52)		N-pyrrolidinaldehyde-2-carboxylic acid	-15.62 (-6.66 \pm 2.68)	
cis-(D-Ala,D-Ala)-2,5-Diketopiperazine	-12.58 (-4.33 \pm 2.84)		Pyridine	-8.98 (-3.10 \pm 1.82)	
2,5-Diketopiperazine	-12.49 (-4.10 \pm 2.69)		2-Pyrrolidinone	-9.22 (-4.63 \pm 1.54)	
3,6-dimethyl-2,5-diketo-1,4-piperazine	-12.59 (-4.87 \pm 2.85)		Quinazoline	-10.49 (-4.00 \pm 2.62)	
N-ethylpyrrolidinone	-14.54 (-6.55 \pm 2.50)		Quinoline	-10.72 (-4.01 \pm 2.81)	
N-ethylimidazole	-9.12 (-3.88 \pm 2.01)		Trans-Dihydroxy-cyclohexane (anti)	-12.17 (-4.89 \pm 2.77)	
Fluorobenzene	-7.73 (-3.62 \pm 1.81)		Tetraline	-11.29 (-5.25 \pm 3.16)	
2,3,4-furantricarboxylic acid	-11.50 (-5.63 \pm 2.67)		Tetrahydro-2-pyrimidinone	-11.35 (-5.07 \pm 1.91)	
Guanine	-12.75 (-4.40 \pm 2.70)		Tetrahydro-quinoline	-12.21 (-4.94 \pm 3.01)	
Quinoline-4,7-diamine	-11.79 (-4.62 \pm 3.21)		Toluene	-8.61 (-4.33 \pm 2.17)	
Quinoline-3,10-diamine	-10.84 (-3.98 \pm 2.80)		Trifluoromethylbenzene	-12.35 (-4.54 \pm 2.77)	
Quinoline-3-E-methylamide-10-amine	-11.63 (-4.76 \pm 2.93)		Uracil	-9.56 (-3.55 \pm 1.67)	

Table 7.18: Calculated interaction energies in kcal·mol⁻¹ (maximum, mean and standard deviation) of molecular probes found inside CURED.

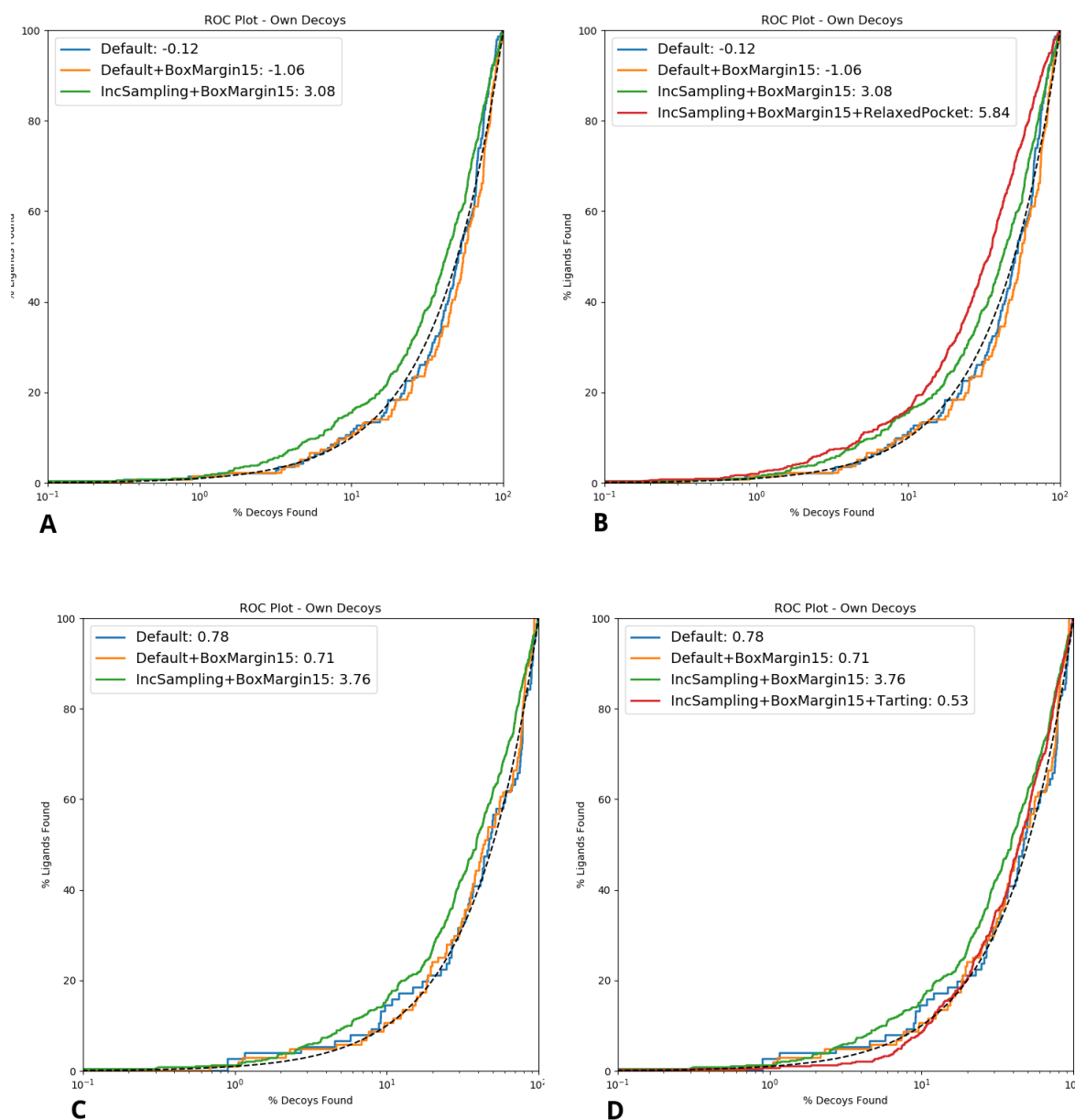


Figure 7.12: ROC plots of molecular enrichment of known actives for prepared SMO Vismodegib (plots A and B) and SANT-1 receptors (plots C and D). In each case, it is demonstrated that increasing the box margin and orientational sampling from defaults enhanced the model's ability to dock and score molecules known to be active against SMO.

TM1	TM2	TM3	TM4	TM5	TM6	TM7	ECL1	ECL2	ECL3
I234	W281	V321		V404	T466	M525	Y207	Q284	D382
		L325		I408	C469	F526	I215	M301	G383
					H470		N219	L303	D384
					Y472		P220		S385
					D473		L221		V386
					F474		F222		S387
					N476		M230		G388
					Q477		H231		I389
					A478		Y233		C390
					E479				F391
					W480				G393
					E481				Y394
					R482				K395
					S483				N396
					F484				Y397
					R485				R400
					V488				
					I509				
					N511				
					R512				
					P513				
					S514				
					L515				
					V517				
					E518				
					K519				
					I520				
					N521				
					L522				
					F523				
					A524				

Table 7.19: Smoothened residues within 8Å of the Vismodegib binding position by helix/loop that were energy minimised by both MOE and CHARMM

	n	min	max	range	\tilde{x}	$\bar{x} (\sigma_{\bar{x}})$
CHARMM	66	0.091	1.019	0.928	0.21	0.40 (0.21)
MOE	66	0.331	1.685	1.685	0.46	0.54 (0.36)

Table 7.20: Descriptive statistics for distance in RMSD of selected residues from X-ray structure of PDB:5L7I of CHARMM-minimised and MOE-minimised receptor structures demonstrating that overall movement of atomic positions in SMO pocket residues was mild when using CHARMM (\tilde{x} =0.40) or MOE(\tilde{x} =0.54).

ENER ENR: Eval#	ENERgy	Delta-E	GRMS		
ENER INTERN:	BONDs	ANGLes	UREY-b	DIHEdrals	IMPRopers
ENER EXTERN:	VDWaals	ELEC	HBONDs	ASP	USER
CHARMM					
ENER> 0	-18341.68878	0.00000	8.92134		
ENER INTERN>	298.29580	1476.16693	0.00000	1259.32198	27.59509
ENER EXTERN>	-4615.67876	-16787.38982	0.00000	0.00000	0.00000
MOE					
ENER> 0	-18197.08623	0.00000	9.39867		
ENER INTERN>	342.27848	1621.19000	0.00000	1286.86442	44.01162
ENER EXTERN>	-4565.37810	-16926.05265	0.00000	0.00000	0.00000

Table 7.21: Calculated receptor energies of receptors after minimisation with CHARMM and MOE of SMO receptors. Despite more overall movement of atom positions in the MOE structure, the CHARMM-minimised structure was overall lower in energy.

Residue code	Atoms affected (AMBER atom types)
VAL270 (VAM)	O (-0.500 → -0.900), H (-0.248 → -0.648)
LEU325 (LEV)	O (-0.500 → -0.900), H (-0.248 → -0.648)
GLY328 (GLZ)	O (-0.500 → -0.900), H (-0.248 → -0.648)
GLY456 (GLZ)	O (-0.500 → -0.900), H (-0.248 → -0.648)
PHE462 (PHF)	O (-0.500 → -0.900), H (-0.248 → -0.648)
THR466 (THS)	O (-0.500 → -0.900), H (-0.248 → -0.648), HOG (-0.310 → -0.710)
ALA524 (ALB)	O (-0.500 → -0.900), HN (-0.248 → -0.648)
MET525 (MEU)	O (-0.500 → -0.900), HN (-0.248 → -0.648)
THR528 (THS)	O (-0.500 → -0.900), H (-0.248 → -0.648), HOG (-0.310 → -0.710)

Table 7.22: Tarted Smoothened residues with atomic partial before and after tarting residues in PDB: 4N4W

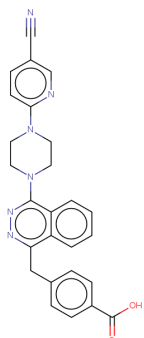
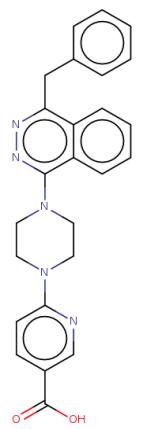
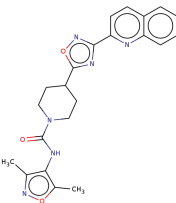
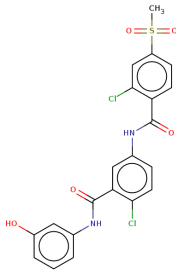
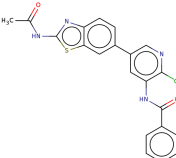
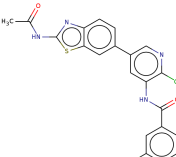
Number	ZINCID	2D Structure	Tarted score	Untarted score
1	CHEMBL497218		-153.07	-4.23
2	CHEMBL523431		-124.25	-1.61
3	CHEMBL182383		-109.37	-28.42
4	CHEMBL312670		-90.91	-5.75
5	CHEMBL360461		-71.71	1.47
6	CHEMBL360461		-69.16	3.57

Table 7.23: Molecules known to be active against SMO that scored high when docked to the tarted receptor but poorly against the untarted receptor and their interaction energies in both docks. This is to demonstrate that the interaction energies of some very polar molecules were scored highly simply due to the increase in electrostatic potential around tarted residues.

	$\bar{x} \pm \sigma_{\bar{x}}$	min	Q1	\tilde{x}	Q3	max
Lacroix, et al	0.29 ± 0.08	0.12	0.23	0.29	0.35	0.55
Round 1	0.52 ± 0.12	0.23	0.44	0.51	0.61	0.85
All actives	0.29 ± 0.10	0.04		0.51		0.71

Table 7.24: Descriptive statistics of ECFP4 Tanimoto similarity of Round 2 molecules with Lacroix et al and Round 1 molecules showing overall low similarity between Round 2 molecules with Round 1 molecules, a prior study using the ZINC12 dataset and all known actives against SMO

	5L7I		4N4W	
	R1 (X-ray)	R2 (relaxed)	R1 (X-ray)	R2 (tarted)
Total Energy	19.54	19.63	17.31	-2.58
Overall: MMFF VdW	144.5	168.77	177.45	185.06
Overall: MMFF Coulomb	-1101.92	-1057.74	-1205.42	-1269.97
Overall: MMFF Torsion	98.55	107.34	90.32	96.68
Prot_ vdw	120.83	144.93	155.12	163.50
Prot: MMFF Torsion	90.01	99.54	83.18	91.71
Lig: MMFF VdW	31.38	34.56	27.62	28.19
Lig: MMFF Torsion	8.54	7.81	7.14	4.97
Prot-Lig: MMFF VdW	-7.72	-10.73	-5.29	-6.63
Prot-Lig: MMFF Coulomb	-28.53	-28.63	-52.63	-33.42
Prot-Lig:.. VdW	-14.75	-17.9	-12.45	-13.86
Prot-Lig:.. Coulomb diel=1.0	-165.1	-135.02	-185.17	-178.37
Prot-Lig.. Protein desolv:(PB)	26.42	25.44	34.50	34.97
Prot-Lig.. Ligand desolv: (PB)	24.93	22.53	22.20	21.26
Overall Prot-Lig Interaction	-17.45	-17.35	-12.15	-12.51

Table 7.25: Mean estimated energies from SZYBKI (kcal.mol⁻¹) for ZINC12 Round 1 vs Round 2 molecules - X-ray vs altered receptor structures

	5L7I		4N4W	
	ZINC12	ZINC15	ZINC12	ZINC15
Total Energy	19.63	-0.16	-2.58	35.81
Overall: MMFF VdW	168.77	165.61	185.06	143.13
Overall: MMFF Coulomb	-1057.74	-1095.80	-1269.97	-855.63
Overall: MMFF Torsion	107.34	100.37	96.68	76.60
Prot_ vdw	144.93	146.05	163.50	140.22
Prot: MMFF Torsion	99.54	95.42	91.71	70.58
Lig: MMFF VdW	34.56	28.67	28.19	18.12
Lig: MMFF Torsion	7.81	4.96	4.97	6.02
Prot-Lig: MMFF VdW	-10.73	-9.12	-6.63	-15.22
Prot-Lig: MMFF Coulomb	-28.63	-52.37	-33.42	-12.48
Prot-Lig:.. VdW	-17.90	-16.92	-13.86	-25.11
Prot-Lig:.. Coulomb diel=1.0	-135.02	-138.42	-178.37	-10.26
Prot-Lig.. Protein desolv:(PB)	25.44	25.75	34.97	22.79
Prot-Lig.. Ligand desolv: (PB)	22.53	21.12	21.26	7.23
Overall Prot-Lig Int	-17.35	-16.46	-12.51	-0.88

Table 7.26: Mean estimated energies from SZYBKI (kcal.mol⁻¹) for Round 2 molecules - ZINC12 vs ZINC15

Chapter 8

Definitions

Abbreviation	Complete definition
Chemistry and biology	
HTS	High-Throughput Screen/ing
TSA	Thermal Shift Assay
vdW	van der Waals
Computational chemistry	
MD	Molecular dynamics simulation
PES	Potential Energy Surace
Pim-1	
ATP	Adenosine Triphosphate
Pim-1	Proviral Integration site for MuLV
Smoothened receptor	
7TM	7 Transmembrane Bundle
CRD	Cysteine-Rich Domain
ECL	Extra-Cellular Loop
GPCR	G-Protein Coupled Receptor
Hh	Hedgehog signal pathway
SANT1	Smoothened ANTagonist 1
SMO	Smoothened receptor
Vismo	Vismodegib
Statistical	
AUC	Area Under the Curve MCC
\tilde{x}	Median
Q1	25% quartile
Q3	75% quartile
\bar{x}	Sample mean
$\sigma_{\bar{x}}$	Sample standard deviation

Chapter 9

Curriculum vitae

Corey John Taylor

corey.taylor@uni-marburg.de

Wehrdaerweg 38a
Marburg 35037

Hessen — Germany
+49 1575 8384534
Date of birth: 12.03.1979
Nationality: Australian

Education

1997–2001	Bachelor of Science , University of Adelaide, South Australia. Majors in organic chemistry and physical & inorganic chemistry
2006–2008	Graduate Diploma in Psychology , Flinders University, South Australia (incomplete; 31.5/36 units complete)
2010–2013	Master of Information Technology , Flinders University, South Australia. Completed coursework in a mixture of advanced programming, software engineering, AI/ML and data modelling topics Thesis title: <i>Statistical models of user interactional context</i> <i>Supervised and unsupervised statistical machine learning approaches to speech recognition with applications in multi-modal fusion. Included AI algorithms in concert with classes/methods in both an open-source automatic speech recognition toolkit (Sphinx4) and general-purpose ML toolkit, the MACHine Learning for Language Toolkit (MALLET)</i>
Nov 2014–Feb 2019	Dr. rer. nat. Kolb Lab, Philipps-Universität Marburg, Hessen, Deutschland. Thesis title: <i>Discovery and development of novel inhibitors for the kinase Pim-1 and G-Protein Coupled Receptor Smoothened</i> <i>The development of novel inhibitors to a diverse range of protein targets using a mixture of chemoinformatics and molecular dynamics methods</i>

Scholarships

2014	Flinders Summer Research Scholarship Project title: <i>Natural Language Processing</i> <i>Development of a AI training tool for health care professionals in the practice of Motivated Interviewing (Miller and Rose, 2009) using an Embodied Conversational Agent (ECA), an animated face which display gestures, facial expressions and speech.</i>
1997–2001	Commonwealth Scientific and Industrial Research Organisation (CSIRO) Molecular Science Cadetship , Melbourne/Adelaide

Research skills

Professional	Grant writing Scientific papers for peer-review Project management as part of a multidisciplinary team
Computational	<i>Extensive experience</i> Chemoinformatics (e.g molecular docking, rescoring, postprocessing) Molecular dynamics simulation (atomistic, MM-PBSA, GIST) Object-Oriented Programming (OOP) and software engineering practices AI/ML and Bayesian approaches (Latent Dirichlet Allocation, Probabilistic Topic Modelling, Conditional Random Field, Bayesian Networks, Dempster-Schafer Networks, HMM) High Performance Computing (e.g. SGE and distributed computing) <i>Some experience</i> Quantum chemistry (DFT, WFT, QTAIM, FMO)
Statistical	Complex linear modelling (GLM/LMM, multiple linear/logistic/ regression analysis, mixed-effects) Survival analysis (Cox regression, Kaplan-Meier, Competing Risks Regression) ANOVA/MANOVA/ANCOVA Structured Equation Modelling and path analysis Parametric and non-parametric methods Time-series analysis (e.g. ARIMA) Cluster analysis (two-step, k-means, PCA) Experimental/sample design, use of and configuration of sampling weights (Jack-knife, BRR, Linearized/Bootstrapping)
Laboratory	FRET-based kinase assay kit

Computer skills

Chemistry	<p>Chemoinformatics/molecular docking</p> <p>UCSF DOCK, OpenEye suite, SEED, OpenBabel, AutodockVINA</p> <p>Molecular dynamics</p> <p>Amber, CPPTRAJ, CHARMM/CHARMMGUI, MDMix</p> <p>Visualisation/modelling</p> <p>Pymol, Chimera, WIP, ChemCraft, VMD, ChemAxon suite, MOE, Molden, JMol, Avogadro, GabEdit, Coot</p> <p>Quantum chemistry</p> <p>Orca, Gaussian, MultiWFN, GAMESS</p>
Programming	<p>OOP</p> <p>Java, C++ and Python</p> <p>Open source machine learning TKs</p> <p>Sphinx4/CMUSphinx, MACHine Learning for LanguagE (MALLET) and OEChem</p> <p>IDEs</p> <p>Eclipse for Java or C++, Netbeans, RStudio</p> <p>Markup/scripting</p> <p>LaTeX, HTML, PHP, Javascript, tch, BASH, Perl</p> <p>Modelling</p> <p>Dia, StarUML</p>
Statistics	Stata, R/R Studio, SPSS, and SAS
Professional	All aspects of the software development life cycle including development frameworks (e.g. Waterfall, Agile), modelling (e.g. UML, class and sequence diagrams, requirements elucidation) testing (e.g. unit tests, TDD, regressions) and version control (e.g. GIT, CVS)
Desktop	All Microsoft-based software (Word, Excel, Powerpoint, Excel, Access, etc.), configuration and use of UNIX-style systems (various Linux distributions including Red Hat, Ubuntu, Slackware, Debian, etc.; Free/OpenBSD; MacOS).
Admin	4+ years experience in configuration, maintenance and administration of a heterogeneous network of 30+ Linux machines and servers at Department of Pharmaceutical Chemistry, Philipps-Universität Marburg

Professional Experience

2012–2014	Royal Australasian College of Surgeons, Senior Research Officer <ul style="list-style-type: none"> • Various epidemiology research projects in conjunction with Cancer Australia on the topic of breast cancer • Research projects with fellows of the Royal Australasian College of Surgeons • Co-author of publications for peer-review • Primary author for National Health and Medical Research Council (NHMRC)/federal government grants
2013–2014	Flinders University, Statistics lab demonstrator <ul style="list-style-type: none"> • Lab demonstrator for biology and computer science students • Topics taught included basic descriptive statistics, t-tests, Analyses of Variance (ANOVA, MANOVA, MANCOVA) and linear regression
2012–2014	Planning and Institutional Performance, University of South Australia, Senior Analyst <ul style="list-style-type: none"> • Responsible for providing research capability for UniSAs business needs • Analysis of large datasets using a range of techniques (time-series, SEM, regression, etc.)
2010–2012	National Centre for Education and Training on Addiction, Statistician <ul style="list-style-type: none"> • Lead member of project team developing a national alcohol knowledgebase, linking multiple research and administrative data sources. • Research papers published in peer-reviewed journals and conference presentations. • Statistical resource for multiple projects • Grant writer for National Health and Medical Research Council (NHMRC)/federal government grants
2011–present	Statistical consultant <ul style="list-style-type: none"> • Statistical advice/analysis provided to various private companies • Hourly-paid work in, for example, high-level analysis of survey data.

Professional Experience (cont.)

2008–2010	<p>Data, Analysis and Research team, Office for Disability and Client Services, Data and Statistical Analyst</p> <ul style="list-style-type: none"> • Perform statistical analyses (quantitative and qualitative) for various projects within the ambit of disability • Lead research projects investigating various issues using numerous data sources at both the state and national level • Act as a representative of national-level data working groups
2008–2008	<p>Clinical Systems Improvement, Mental Health Unit, Department of Health, Adelaide, SA, Senior research officer (6-month contract)</p> <ul style="list-style-type: none"> • Lead member on a project to develop an evidence-based standardised risk assessment tool for mental health patients in emergency and community settings. • Investigated decline in reporting of community suicides and process for improvement of reporting • Required liaison, consultation and negotiation with senior management and front-line clinicians throughout the Department of Health to achieve the aims of the project and ensure wide consultation
2004–2008	<p>Intelligence Section, Sexual Crime Investigation Branch, South Australia Police., Intelligence Analyst</p> <ul style="list-style-type: none"> • Real-time operational response & intelligence capability to sexual crime which met SCIBs criteria • Intelligence capability to long-term child abuse investigations • Maintain awareness of trends in sexual/violent crime and inform decision-makers accordingly. • Responsible for high-level intelligence reports to senior management and other members of SAPol up to the level of Commissioner • Responsible for maintaining strong professional relationships with multiple and diverse stakeholders, mediating negotiations between them and acting as a point of contact for briefs I was the lead operator on

Professional Experience (cont.)

2003–2004	<p>Intelligence Analysis Section, State Intelligence Branch, South Australia Police., Intelligence Analyst</p> <ul style="list-style-type: none"> • Intelligence function to inform and support SAPol operations/investigation within the ambit of vehicular crime, robbery, drugs and repeat offenders • Trained in anti-terrorism operational and intelligence methodologies • Responsible for maintaining strong professional relationships with multiple and diverse stakeholders, mediating negotiations between them and acting as a point of contact for briefs I was the lead operator on
2000–2004	<p>Various roles Research Officer, Aboriginal Housing Authority; Graduate Officer, Department of Human Services; Hourly-Paid Lecturer, Gilles Plains Tafe; Tutor, University of Adelaide</p>
1997–2000	<p>IGF Agonists and Antagonists Group, CSIRO Division of Molecular Engineering, Adelaide and Melbourne, Cadet officer</p> <ul style="list-style-type: none"> • Responsible for support of research staff in performing various functions • Duties included sterile cell culturing, ELISA Immunoassays, various spectrophotometric techniques (Atomic Absorption Spectroscopy, High-Performance Liquid Chromatography, etc.) and chemical analysis using standardised analysis techniques

Publications

As at 19.11.18:

Number of peer-reviewed journal publications: **11**
h-index: **7**

See my Google scholar profile for the most up-to-date list of publications:

<http://scholar.google.com.au/citations?user=IjR7hksAAAAJ&hl=en>

Most recent publications:

- 2015 C Taylor, R Leibbrandt, and D. Powers. Anywhere but here: Enron's emails in the midst of a crisis. *International Conference on Agents and Artificial Intelligence (ICAART 2015)*, 8, 2015
- 2014 J Azzopardi, D Walsh, C Chong, and C Taylor. Surgical treatment for women with breast cancer in relation to socioeconomic and insurance status*. *The breast journal*, 20(1):3-8, 2014, 2014
- 2013 D Roder, H Zorbas, J Kollias, C Pyke, D Walters, I Campbell, C Taylor, and F Webster. Factors predictive of treatment by australian breast surgeons of invasive female breast cancer by mastectomy rather than breast conserving surgery. *Asian Pacific Journal of Cancer Prevention* 14 (1), 539-545, 2013
- D Roder, H Zorbas, J Kollias, C Pyke, D Walters, I Campbell, C Taylor, and F Webster. Factors predictive of immediate breast reconstruction following mastectomy for invasive breast cancer in australia. *The Breast* 22 (6), 1220-1225, 2013
- C Chong, D Waters, P deSilva, C Taylor, A Spillane, J Kollias, C Pyke, and I Campbell. Subsequent axillary surgery after sentinel lymph node biopsy: Results from the breastsurganz quality audit 2006-2010. *The Breast* 22 (6), 1215-1219, 2013
- D Roder, H Zorbas, J Kollias, C Pyke, D Walters, I Campbell, C Taylor, and F Webster. Risk factors for poorer breast cancer outcomes in residents of remote areas of australia. *Asian Pacific Journal of Cancer Prevention* 14 (1), 547-552, 2013

Conference papers and presentations

- | | |
|------|--|
| 2016 | C Taylor, C Siefker, H Rimmer, W Diederich, G Klebe, and P Kolb. Fragment-based development of selective kinase inhibitors. In <i>SynChemBio Symposium, 16. February 2016, Rauischholzhausen</i> , 2016 |
| 2014 | D Roder, J Kollias, I Campbell, D Walters, and C Taylor. Factors predictive of immediate breast reconstruction following mastectomy for invasive breast cancer in australia circa 1998-2010. In <i>Royal Australasian College of Surgeons (RACS) Annual Scientific Congress and the Australian and New Zealand College of Anesthetists (ANZCA) Annual Scientific Meeting, 5th-9th of May, Sands Expo and Convention Center Marina Bay Sands, Singapore</i> , 2014

A Vaskas, J Kollias, and C Taylor. Sentinel lymph node biopsy should not be performed for screen-detected ductal carcinoma in situ undergoing breast conserving surgery. In <i>Royal Australasian College of Surgeons (RACS) Annual Scientific Congress and the Australian and New Zealand College of Anesthetists (ANZCA) Annual Scientific Meeting, 5th-9th of May, Sands Expo and Convention Center Marina Bay Sands, Singapore</i> , 2014 |
| 2012 | J Azzopardi, D Walsh, and C Taylor. Surgical treatment for women with breast cancer in relation to remoteness, socioeconomic disadvantage and insurance. In <i>General Surgeons Annual Scientific Meeting, 21-23rd of Sept, Wrest Pt Hobart, Tas</i> , 2012 |
| 2010 | A Roche, K Pidd, and C Taylor. A strategy to improve usability and accessibility of alcohol data through data linkage. In <i>PHAA Conference, 27-29th of Sept 2010, Adelaide</i> , 2010 |

Conference posters

- | | |
|------|---|
| 2018 | C Taylor, C Siefker, L Heyder, F Chevillard, W Diederich, G Klebe, and Kolb P. Always look twice - turning some good fortune into ligands for the kinase pim-1. In <i>German Chemoinformatics Conference, 11th-13th of Nov, Mainz, Germany</i> , 2018 |
| 2017 | C Taylor, C Siefker, F Chevillard, H Rimmer, W Diederich, G Klebe, and Kolb P. Fragment-based development of selective kinase inhibitors - pim-1 and pka. In <i>Spring School in Computational Chemistry CSC, Mar 12th-16th Espoo, Helsinki, Finland</i> , 2017 |

Courses completed

- | | |
|---------|--|
| 2017 | Spring School in Computational Chemistry CSC, Mar 12th-16th Espoo, Helsinki, Finland |
| 2015 | EUROPIN Summer School on Drug Design, Sep 20th-25th, Vienna, Austria |
| 2006 | Criminal intelligence analysis course, Sep 4th-16th, South Australia Police Academy, Adelaide, South Australia |
| Various | Multiple professional courses in analysis (e.g. SPSS), professional writing, public speaking |

Referees

- 1 Prof. Dr. Peter Kolb

Heisenberg Professor of Pharmaceutical Chemistry, Kolb Lab, Philipps-Universität Marburg, Marburg, Deutschland

Relationship: PhD supervisor

☎+49 6421 28 25908 ✉peter.kolb@uni-marburg.de
- 2 Prof. Dr. Gerhard Klebe

Professor for Pharmaceutical Chemistry, Philipps-Universität Marburg, Marburg, Deutschland

Relationship: PhD co-supervisor

☎+49 6421 28 25908 ✉peter.kolb@uni-marburg.de
- 3 Prof. David Roder

Chair of Cancer Epidemiology and Population Health, School of Health Sciences, University of South Australia

Relationship: Research collaborator

☎+61 8 8302 2640 ✉David.Roder@unisa.edu.au
- 4 Dr. Richard Leibbrandt

Research Software Development Assistant, College of Medicine and Public Health, Flinders University

Relationship: Master thesis supervisor

☎+61 8 8201 3659 ✉richard.leibbrandt@flinders.edu.au
- 5 Mrs Katherine Economides

Manager Breast Audit, Audit and Academic Surgery Division, Royal Australasian College of Surgeons

Relationship: Former manager

☎+61 8219 0912 ✉katherine.economides@surgeons.org

Signed:

Date:

M U S T E R

E R K L Ä R U N G

Ich versichere, dass ich meine Dissertation

„Discovery and development of novel inhibitors for the kinase Pim-1 and G-Protein Coupled Receptor Smoothened“

selbständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen bedient habe. Alle vollständig oder sinngemäß übernommenen Zitate sind als solche gekennzeichnet.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den **09.04.2019**.

Corey Taylor



.....
(Unterschrift mit Vor- und Zuname)